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

Volume Author/Editor: Victor Zarnowitz

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-97890-7

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

Conference Date: n/a

Publication Date: January 1992

Chapter Title: An Analysis of Annual and Multiperiod Quarterly Aggregate Forecasts

Chapter Author: Victor Zarnowitz

Chapter URL: http://www.nber.org/chapters/c10385

Chapter pages in book: (p. 414 - 443)

14 An Analysis of Annual and Multiperiod Quarterly Aggregate Forecasts

14.1 On Some Uses and Limitations of Forecast Data

How and how well economists forecast, and how much their predictions help or hurt public and private decision making, are matters that ought to receive much attention from the profession. This is so not only because of their direct interest to the authors, users, and critics of the forecasts, but also because of their intrinsic but less evident academic interest. What is the practical applicability of economic analysis in this critical area? What is the quality of foresight and counsel that can be expected of responsible economists? These are broad questions which are not easy to answer, but they are basic and surely deserve to be tackled. This requires that we systematically confront forecasts as indications of how economists ex ante thought events were likely to unfold with ex post knowledge of what actually did happen and how. The aims, from the least to the most ambitious, are (1) to measure forecast errors, (2) to explain them, and (3) to learn how to reduce them in the future.

Success in forecasting may be occasional and fortuitous or intuitive, but progress in forecasting, to the extent it is possible, can only come from advances of science, not art or chance. It presupposes that sufficiently important and persistent regularities in economic processes and relationships exist and

Reprinted from the *Journal of Business* 52 (January 1979): 1–33.

The author is grateful to Dennis Bushe, Thomas Kutzen, and Robert Osterlund for assistance with research and preparation of this paper; to Stephen K. McNees and Linda G. Martin for comments and aid with the data; and to Otto Eckstein, Geoffrey H. Moore, and Arthur M. Okun for criticisms and suggestions based on an earlier draft. This report was supported by grants from the National Science Foundation and the Earhart Foundation to the National Bureau of Economic Research and by a grant from the Center for the Management of Public and Nonprofit Enterprise, Graduate School of Business, University of Chicago.

are properly identified and used. Learning processes are involved, which can be time-consuming and discontinuous, reflecting in part the shifts and discontinuities in the economic change itself, in part the inadequacies of measurement and analysis.

Data on economic forecasts generally cover short time periods. Long time series of consistent predictions simply do not exist. Few if any forecast sets are fully identified according to the many aspects and dimensions that matter (source, target, timing, assumptions, data, models, and methods used), so that it is often difficult to determine what constitutes a suitable "sample" of forecasts of a given type. Moreover, few forecasters leave their models and techniques unchanged for long as they seek improvements and try to adapt to new developments in the economy. Hence, a particular forecaster's past record is often a highly uncertain basis for inferences about future performance.

Even more hazardous, if not irresponsible, are attempts to grade forecasters on the evidence of how well they predicted change in a particular short period, say, a year or a few years. Clearly, on any single occasion some forecasters will be ahead of others by sheer chance or for some idiosyncratic reasons. Strong evidence of significant and stable differences over time is required to rank the forecasting individuals, groups, or models with a modicum of confidence, and such evidence is essentially lacking (Zarnowitz 1967, 1971; Christ 1975; McNees 1975).

The proliferation in recent years of multiperiod quarterly macroforecasts offers no substitute for long historical series. These are rich data containing much interesting material that certainly deserves to be carefully recorded and analyzed. However, such forecasts, and their errors, tend to be internally correlated in at least two ways: (1) serially, within each sequence made from a given base period, and (2) across the successive sequences, which overlap and thus refer partially to the same target period. Each multiperiod forecast is a joint product of the common information, technique, and judgment used, and each depends on previous forecasts of which it is to some extent a revision. Thus, errors in the data, models, procedures, and judgments, autocorrelated disturbances, and certain types of distributed lags are all likely to induce interdependencies within and between the multiperiod forecasts. The resulting complex correlation structures resist estimation, given the small samples of comparable predictions from any given source. Consequently, measures of average accuracy, bias, etc., calculated from such samples are difficult to interpret from the viewpoint of statistical inference (Spivey and Wrobleski 1978).

Two conclusions are surely valid. First, small-sample studies of forecasts are still needed and can be instructive, but their limitations must be recognized. Second, it is necessary to compile and examine forecast records extending as far back in time as possible, so as to gain information, take a longer view of forecasting behavior and performance, and place the short records of recent predictions in a proper perspective. Historical data on post–World War

II forecasts assembled in the 1960s by the National Bureau of Economic Research (NBER) provide a good base here, which I was able to partially extend and update. Some preliminary results for annual forecasts of three variables are reported below.

14.2 The Record of Annual GNP Forecasts since 1947

In the early post–World War II period, most forecasts were made near the end of the calendar year for the next year and most referred to GNP in current dollars. The evidence we have on such forecasts goes back to 1947 but is quite fragmentary for the late 1940s and early 1950s.

The period of transition from the war economy witnessed the largest errors on record in the GNP forecasts. Even after the 1945–46 predictions were shown to have greatly underestimated the then-prevailing levels of economic activity (L. R. Klein 1946), expectations of a business slump stubbornly persisted. One small, reputable group of private forecasters came up with an average prediction for 1947 of a 6% decline in GNP, whereas the actual change turned out to be a rise of about 11%. For 1948 the group predicted a fractional decline, but GNP instead advanced again at much the same surprisingly high rate. The failure of forecasts during these years was widespread, with but a few partial exceptions; the developments of the time could not have been predicted well with estimates based on data and relationships for the 1930s and false analogies with the early post–World War I period, and were not. When a recession finally came late in 1948, it proved shorter than many had expected. A "consensus forecast" by more than 30 respondents polled in December 1948 anticipated well the decline of nearly 2% in GNP during 1949, but a year later the same group was wide of the mark in predicting a drop of 3:5%, whereas GNP actually staged a strong comeback in 1950 with a rise exceeding 10% (though it should be noted that the latter resulted in part from the onset of the Korean War in June 1950).

The evidence for the period 1953–76 is summarized in table 14.1 in terms of comparisons between the predicted and the actual annual percentage changes. It is generally instructive to analyze forecast errors in terms of levels, absolute changes, and percentage changes, but if a choice must be made for succinctness, there are several good reasons for using percentage changes where technically appropriate, particularly for variables with strong trends. (1) What is predicted in the first place is change from the last known or estimated level, and percentage changes often depend less on the levels and are more stable and comparable over time than dollar changes. (2) Percentage change forecasts are apt to be less affected by data revisions. (3) Some important measures of predictive performance, such as correlations with actual values, are much more meaningful for change forecasts than for level forecasts. (4) It is the rates of growth in economic aggregates (income, output, prices) that are of main interest to analysts and policymakers.

The forecasts are made late in the year $t-1$ or, in a few cases, very early in the target year t; typically, the forecasters know the official estimates for the first three quarters but not for the last quarter of the year $t-1$. The actual changes used to compute the errors are based on the first official estimates for the year t published early in the following year $(t+1)$. These are provisional values which are themselves partly near-term predictions, and subsequent revisions indicate that the errors in the early data are by no means negligible (cols. 10–11).[1] On the average, without regard to sign, these revisions are about one-third the size of the forecast errors (lines 5–8). The errors are computed by subtracting the actual from the predicted (or estimated) changes, and they are predominantly negative, which shows that both forecasts and the provisional figures strongly tend to understate the changes in GNP (lines 9–12). By far most of these underestimated changes are increases (for a review of similar findings of earlier studies, see Zarnowitz 1972b).

Table 14.1 discloses a substantial correspondence between the forecasts and the realizations. The predicted changes approximate the actual ones well in each period covered, the averages of the former being generally less than 1 percentage point smaller than the averages of the latter (lines 1–4). Where the mean actual changes increased (as from 5% per year in 1956–63 to 8% per year in 1963–76), so did the mean predicted changes; moreover, the discrepancies between the two diminished in the latter years. The forecasts are in all cases considerably more accurate than the naive model which assumes that next year's percentage change will be the same as that of the previous year and more accurate than the—somewhat less naive—trend extrapolation model, which projects the average percentage change of the 4 previous years. Collectively, the mean absolute error (MAE) of forecasts is less than half that of the first naive model (lines 5–8, col. 8), and the ratio of the two declines from 0.47 in 1956–63 to 0.43 in 1963–76 and 0.34 in 1969–76. The corresponding ratios for comparisons of the forecasts with the 4-year moving average ("trend") extrapolations (col. 9) are 0.84, 0.56, and 0.44.

The average error measures are important, but they fall far short of telling the whole story. Measures of correlation (which unfortunately are often omitted from forecast evaluations) are needed to show how well the predicted changes have tracked the actual changes over time. The r^2 coefficients for the forecasts covered in table 14.1 are all positive and significant, generally exceeding .5 and, for the more recent periods, averaging .7 or higher (lines 13–

1. In some evaluations, forecasts are adjusted for the differences between the forecaster's own estimates of the base (current actual value) and the revised base figures (see McNees 1975, pp. 7–9, and Eckstein 1978, p. 320, for arguments in favor of that practice). My own view is that it is more instructive to study the base revisions separately as one of the sources contributing to the total error, that is, to compute and compare both the level errors unadjusted and the absolute change errors adjusted for base revisions. For percentage change errors, which to save space are alone used here, I agree with McNees (1975, p. 9) that "it is not clear what adjustments for data revisions would be appropriate or even possible." Preliminary calculations indicate that the effects of revisions on errors computed in this form are small and not systematic.

Table 14.1 Summary of Measures of Error for Annual Predictions of Percentage Change in GNP, 1953–76

Line	Period and No. of Years Covered	Livingston Survey, Mean Forecast[a] (1)	Selected Private Forecasts, Mean[b] (2)	N.Y. Fore-casters Club, Mean Forecast[c] (3)	ASA-NBER Survey, Median Forecast[d] (4)	*Economic Report of the President*[e] (5)	Michigan Model[f] (6)	Wharton Model[g] (7)	Extrapolations: Last Change[h] (8)	Extrapolations: Average Change[i] (9)	Actual: Prelim-inary[j] (10)	Actual: Revised[k] (11)
						Mean Absolute Percentage Change, Predicted and Actual						
1	1953–76 (24)	5.7	6.0	. . .	. . .	. . .	. . .	. . .	. . .	. . .	6.6	6.9
2	1956–63 (8)	4.3	4.7	4.0	. . .	. . .	. . .	. . .	. . .	. . .	5.0	5.3
3	1963–76 (14)	7.1	7.2	. . .	. . .	7.6	7.2	7.6	. . .	. . .	7.9	8.2
4	1969–76 (8)	8.1	8.0	. . .	8.1	8.6	8.3	8.2	. . .	. . .	8.4	8.8
						Mean Absolute Error, in Percentage Points						
5	1953–76 (24)	1.6	1.2	. . .	. . .	. . .	. . .	. . .	3.1	2.3	.5	. . .
6	1956–63 (8)	1.7	1.4	1.7	. . .	. . .	. . .	. . .	3.4	1.9	.5	. . .
7	1963–76 (14)	1.1	.9	. . .	. . .	.9	1.3	.8	2.3	1.8	.3	. . .
8	1969–76 (8)	1.0	.6	. . .	1.0	.8	1.0	.9	2.6	2.0	.3	. . .
						Mean Error, in Percentage Points						
9	1953–76 (24)	−1.0	−.7	. . .	. . .	. . .	. . .	. . .	−.3	−.1	−.5	. . .
10	1956–63 (8)	−.4	−.1	−.8	. . .	. . .	. . .	. . .	−.2	−.4	−.3	. . .
11	1963–76 (14)	−.8	−.6	. . .	. . .	−.2	−.5	−.3	−.3	−.6	−.3	. . .
12	1969–76 (8)	−.3	−.4	. . .	−.3	.2	−.1	−.2	−.3	−.5	−.3	. . .
						Squared Correlation (r^2); between Predicted and Actual Change						
13	1953–76 (24)	.739	.792	. . .	. . .	. . .	. . .	. . .	.012	.054	. . .	. . .
14	1956–63 (8)	.634	.497	.563	. . .	. . .	. . .	. . .	.155[l]	.038	. . .	. . .
15	1963–76 (14)	.717	.791	. . .	. . .	.752	.603	.689	.006	.079	. . .	. . .
16	1969–76 (8)	.780	.875	. . .	.768	.834	.746	.669	.002	.000	. . .	. . .

[a]Based on surveys conducted by Joseph A. Livingston, syndicated columnist with the *Philadelphia Inquirer*, and published in the *Philadelphia Bulletin* and *American Banker*. Of the semiannual surveys, only the end-of-year ones are used here; they typically cover answers to a questionnaire mailed in November and appear in a "Business Outlook" column late in December. The participants in these surveys, listed at the end of the *Bulletin* columns, varied in number between 44 and 62.

[b]Average of end-of-year annual GNP forecasts from the following sources: (1) *Fortune* magazine ("Business Roundup"); (2) Harris Trust and Savings Bank; (3) IBM Economic Research Department; (4) National Securities and Research Corporation; (5) NICB (now Conference Board) Economic Forum; (6) Robert W. Paterson, University of Missouri; (7) Prudential Insurance Company of America; (8) UCLA Business Forecasting Project. The earliest of these predictions were made in October; the latest in January. Most but not all of the forecasts in each of these eight sets are available in published form; those for the period ending in 1969 were analyzed in NBER studies of economic forecasting (Zarnowitz 1967, 1972b, 1974).

[c]Of the semiannual forecasts of this group, only the end-of-year ones are included. The group mean forecasts used here cover individual predictions varying in number between 31 and 39. These data, too, were analyzed in NBER studies (see ref. in n. *b*), but no forecasts were collected for the period after 1963. The predictions for 1956–58 were made in October; those for 1959–63 in December.

[d]Source: Quarterly releases by the American Statistical Association and the National Bureau of Economic Research, published in the ASA *AmStat News* and the NBER *Explorations in Economic Research*. Median forecasts from the November surveys only are used. The membership in these surveys varied between 45 and 84. See Zarnowitz (1969a) and Su and Su 1975.

[e]Forecasts by the Council of Economic Advisers (CEA) as stated in the *Economic Report* (usually as the midpoint in a relatively narrow range). As a rule, the *Economic Report* appears in January. For some earlier studies of these forecasts, see Moore 1969a, 1977d; Zarnowitz 1972b; Fellner 1976, pp. 118–24; McNees 1977.

[f]Published ex ante forecasts from the Research Seminar in Quantitative Economics (RSQE) of the University of Michigan. Based on several working models (see Suits 1962; Hymans and Shapiro 1970, 1974). The forecasts are those released in connection with the University of Michigan annual "Conference on the Economic Outlook," occurring usually in November (in 1974 and 1975, December).

[g]Source: *Wharton Economic Newsletter*, Econometric Forecasting Unit, Wharton School of Finance and Commerce, University of Pennsylvania. Forecasts based on a series of versions of Wharton models (see Evans and Klein 1967; Evans, Klein, and Saito 1972b; McCarthy 1972; Duggal et al. 1974). The forecasts here covered are dated in November or (as in 1971 and 1973–75) in December.

[h]Assumes that next year's percentage change will be the same as that of the previous year. The actual changes used are those based on the preliminary estimates explained in n. *j* below.

[i]Assumes that next year's percentage change will be the same as the average percentage change in the 4 previous years. On the actual changes used, see n. *j*.

[j]Based on the first official estimates following the year for which the forecast was made.

[k]Based on current data taken from U.S. Department of Commerce, Bureau of Economic Analysis, *Handbook of Cyclical Indicators, A Supplement to the Business Conditions Digest* (May 1977).

[l]r is negative.

16, cols. 1–7). In contrast, the corresponding coefficients for the extrapolations (cols. 8 and 9) are zero or near zero (where larger, r is negative). These results mean that the extrapolations can account for little or none of the variation in year-to-year changes in GNP, while the forecasts have captured most of this variation.

Because sufficiently long and consistent annual time-series data for GNP are not available, no attempt was made here to test the forecasts against higher standards provided by more effective extrapolation methods such as the autoregressive integrated moving-average (ARIMA) models. However, recent comparisons of quarterly forecasts with such models show the forecasts to be on the average more accurate (Hirsch, Grimm, and Narasimham 1974; Christ 1975; Spivey and Wrobleski 1978), and I would expect this to be a fortiori true for the annual forecasts and particularly with respect to the correlations with the actual values.

The evidence supports the conclusion that the end-of-year forecasts of current-dollar GNP next year had a reasonably satisfactory record of accuracy since 1953. Indeed, in comparisons with earlier forecasts (Sapir 1949; Okun 1959; Zarnowitz 1967), that record improved considerably in the 1960s and even in the 1970s, a turbulent period presumed to have been particularly difficult to forecast.

It must be noted that our collection is certainly no random sample, including as it does the official administration forecasts and several of the most reputable and influential sets of private predictions by business and academic economists (see notes to table 14.1). It is also true that our data and measures have some shortcomings that must not be overlooked. In particular, the estimates of the current position (ECP) which the forecaster actually used as the starting point or base are not always reported. In some cases, therefore, the base values had to be imputed, which was done using data as of the (precise or approximated) date of the forecast plus such information as was available on how the forecaster derived his ECPs on other occasions. The imputations, even though carefully made, undoubtedly contain some errors. However, these errors are definitely not such as to invalidate the broad conclusions of this paper.[2]

It is difficult to draw more detailed inferences concerning the relative accuracy of the different forecast sets covered from these results. One reason is that the forecasts differ appreciably with regard to their precise dates, and it is known from previous research that later predictions have a significant advantage over earlier ones (Zarnowitz 1967; McNees 1975). It is apparent, however, that the average error and correlation measures do not show large, consistent differences among these forecast sets. This is in agreement with earlier

2. Other possible errors, also not critical, might arise from the fact that some of our forecasts, lacking directly reported annual predictions, are averages of forecasts for shorter periods within the coming year. This could cause some deviations from the span or target period intended by the forecasters (Carlson 1977).

findings which together strongly suggest that the search for a consistently superior forecaster is about as promising as the search for the philosophers' stone (Zarnowitz 1971; McNees 1973, 1975, 1976; Christ 1975).

A few further observations seem warranted. Although the forecasters included differ in many respects, even a detailed inspection reveals few sharp contrasts between their predictions for the same years. Of course, competent forecasters use common data and techniques, regularly interact, and are often similarly influenced by recent events and current attitudes and ways of thinking. The genuine ex ante forecasts here considered are all to a large extent "judgmental." Large doses of judgment enter, mostly helpfully, the forecasts derived with the aid of econometric models (see, e.g., Haitovsky, Treyz, and Su 1974). This could well tend to reduce the dispersion among the corresponding predictions of this type; there is indeed some evidence that errors of ex ante forecasts with econometric models vary less than errors of ex post forecasts made without judgmental adjustments (Christ 1975). At the same time, many so-called judgmental forecasters also use some more or less explicit econometric equations or models, "outside" or "own" (Zarnowitz 1971; Su and Su 1975). While published forecasts by ranking practitioners are often developed with particular skill and care, group average forecasts greatly benefit over time from cancellations of individual errors of opposite sign (Zarnowitz 1967, 1972b). At any given time, the deviations between corresponding forecasts from different sources are likely to be reduced by the working of these balancing factors. Thus, it is not surprising that forecasts for the same variable and target period tend to be similar. Indeed, the correlations between pairs of the forecast sets included in table 14.1, computed for the four periods distinguished therein, are significantly higher than the correlations between predictions and realizations recorded on lines 13–16. The r^2 coefficients for eight pairs of the predicted percentage change series all exceed .8, and some are considerably higher.

Of the 110 observations in our seven forecast sets, about 64% are underestimates of change and 34% are overestimates. By far most of the latter refer to years marked by economic recessions (1954, 1960, 1970, 1974) or slowdowns (1962, 1967). The provisional GNP values show but two year-to-year declines in the period covered in table 14.1: in 1954, which the forecasts overstated, and in 1958, which the forecasts missed (accounting for the only turning-point errors in this sample). Thus underestimation was limited to the increases in GNP; moreover, it was most pronounced when the increases were particularly large, as in 1953, 1955–56, 1965–66, 1968–69, and 1973.

These results suggest the presence of "systematic" errors, but not in the sense of a bias that could have been readily escaped or corrected in advance. It seems difficult to discount them as merely another manifestation of the familiar tendency of forecasts to underestimate the observed changes (which, for series with random elements, is a property of even unbiased and efficient forecasts [Mincer and Zarnowitz 1969; Hatanaka 1975]). What is underestimated here is the average annual rate of growth in a series which, as properly recognized by the forecasters, is trend dominated and seldom declines from year to year. This outcome can be traced to the forecasters' tardy recognition of high-growth phases ("booms") and, increasingly, of inflation speedups, but it was also mitigated by their even tardier recognition of business recessions and slowdowns. Such movements are recurrent and not purely random; they have important, detectable regularities as shown by historical studies of business cycles; but they are also nonperiodic and indeed vary a great deal over time, so their predictability remains very limited. In any event, simple "learning from past errors" would not have been of much use here as the errors of these forecasts generally have zero or very low autocorrelations.

14.3 Annual Forecasts of Real GNP and the Price Level

It is difficult to obtain and verify consistent forecasts of GNP in constant dollars and the implicit price deflator (IPD) that would cover more than just the most recent period. Few business forecasters in the 1950s and 1960s made systematic efforts to decompose their predictions of current-dollar GNP into quantity and price elements. Of the forecasters with econometric models, who paid more attention to real GNP, only two (the University of Michigan and Wharton Econometric Forecasting Associates) have longer records.[3]

Table 14.2 shows that the predicted changes in real GNP, taken without regard to sign, differed from the actual changes by less than 1 percentage point on the average (lines 1–3). The predicted changes tend to be smaller than the actual ones, except for the Council of Economic Advisers (CEA) forecasts (col. 3), where the reverse obtains. The MAEs of the forecasts average a little over four tenths of those of the simple last-change extrapolations in 1959–67 and 1962–76, about one third in 1969–76 (lines 4–6, col. 6). Comparisons with extrapolations of the average percentage change of the 4 previous years give very similar results, except for 1959–67, where the forecast errors average about two thirds of the extrapolation errors (lines 4–6, col. 7).

Correlations between the predicted and actual changes are all significantly positive, and they too suggest some improvement in recent years: the r^2 coefficients for 1969–76 are higher than those for the earlier and longer periods (lines 10–12). It is interesting to observe that all but one of them exceed the corresponding coefficients for current-dollar GNP forecasts, particularly so for the predictions with the Wharton models and the ASA-NBER group medians (from quarterly surveys regularly conducted by the American Statistical Association and evaluated by the National Bureau of Economic Research)

3. Some of the econometric forecasts were released at more than one date near the end of the year, and in more than one version depending on the data used or policy assumptions made. In all but a few doubtful instances where somewhat arbitrary decisions had to be made, the forecasts chosen are those preferred by the forecaster or, lacking stated preferences, those which embodied assumptions most common to the forecasts made at the time.

Table 14.2 **Summary Measures of Error for Annual Predictions of Percentage Changes in Real GNP, 1959–76**

Line	Period and No. of Years Covered	Selected Private Forecasts, Mean[a] (1)	ASA-NBER Survey, Median Forecast (2)	*Economic Report of the President*[b] (3)	Michigan Model (4)	Wharton Model (5)	Extrapolations: Last Change (6)	Extrapolations: Average Change (7)	Actual: Preliminary (8)	Actual: Revised (9)
					Mean Absolute Percentage Change, Predicted and Actual					
1	1959–67 (9)	4.1	. . .	. . .	3.7	. . .	. . .	. . .	4.3(5.0)[c]	4.5
2	1967–76 (15)	. . .	. . .	4.4	3.8	. . .	. . .	. . .	4.1	4.05
3	1969–76 (8)	. . .	3.3	4.0	3.5	3.2	. . .	. . .	3.6	3.3
					Mean Absolute Error, at Percentage Points					
4	1959–67 (9)	1.3	. . .	. . .	1.0	. . .	2.7	1.7	.5	. . .
5	1962–76 (15)	. . .	. . .	1.1	1.4	. . .	2.8	2.6	.4	. . .
6	1969–76 (8)	. . .	1.0	1.2	1.6	.9	3.7	3.6	.3	. . .
					Mean Error, in Percentage Points					
7	1959–67 (9)	−.9	. . .	. . .	−.5	. . .	−.6	−1.1	−.2	. . .
8	1962–76 (15)	. . .	. . .	.6	.2	. . .	−.3	.1	−.1	. . .
9	1969–76 (8)	. . .	.7	.8	.8	.5	−.15	.7	.06	. . .
					Squared Correlation (r^2) between Predicted and Actual Change					
10	1959–67 (9)	.445	. . .	. . .	.531	. . .	.306[d]	.000	. . .	. . .
11	1962–76 (15)	. . .	. . .	.775	.617	. . .	.012	.049[d]	. . .	. . .
12	1969–76 (8)	. . .	.936	.857	.709	.941	.001	.320[d]	. . .	. . .

Note: For sources and explanations of the data used in cols. 2–9, see nn. *d–k*, respectively, in table 14.1.

[a]Average of end-of-year annual forecasts of real GNP inferred from the forecasts of current-dollar GNP, the consumer price index (CPI), and the wholesale price index (WPI) from the following sources: (1) Harris Trust and Savings Bank; (2) National Securities and Research Corporation; (3) NICB (Conference Board) Economic Forum; (4) Robert W. Paterson, University of Missouri; (5) UCLA Business Forecasting Project. These forecasts were obtained by dividing the forecasts of GNP, as reported in current dollars, by the composite price-level forecasts. The latter are weighted sums of the reported forecasts of CPI and WPI, the weights being .647 and .353, respectively (the first of these proportions represents the average ratio of consumption expenditures to GNP in the period 1953–64). For further detail and analysis of the individual forecasts in this set, see Zarnowitz 1969b; see also table 14.1, n. *b*.

[b]The forecasts for 1962, 1963, 1965, and 1968 must be inferred from statements in the *Report;* they are confirmed by the CEA as approximately correct, though not in all cases precisely correct (Moore 1977d). The other forecasts are all based on figures given in the *Report* and so are fully verified.

[c]The figure in parentheses is based on preliminary GNP figures deflated by weighted averages of the corresponding data for CPI and WPI (with weights as given in n. a. above). This series of "actual" values is comparable to the forecasts used in col. 1 only.

[d]r is negative.

(compare table 14.2, line 12, and table 14.1, line 16). In contrast to the reasonably high correlations for the forecasts proper, those for the extrapolations (cols. 6–7) are here again extremely low or negative.

These summary measures, then, present the annual forecasts of real GNP in a generally favorable light. However, the accuracy of these forecasts varied greatly in different years, which at times seriously impaired their usefulness, and this does not show up in the summary. As suggested by the averages with regard to sign (lines 7–9), the usual tendency of forecasts to underestimate changes prevailed in the first half of the period 1959–76 but not in the second half. Actually, the errors varied considerably in each subperiod, primarily reflecting cyclical change and in particular the disturbing effects of missed downturns. Real GNP turned down in 1954, 1958, 1970, and 1974, but of the 10 predictions for these years which are available, 8 specified continued rises and only 2 succeeded in signaling declines. Again, and not surprisingly, nearly all of the significantly large overestimation errors refer to the years during which national output grew at relatively low or decreasing rates, and most of the larger underestimation errors refer to the years of high real-growth rates.

It is of considerable interest to note that the turning-point errors are much larger than other errors (on the average about 2½–3 times larger, for all forecasts in this collection). Thus, even though relatively few, these directional errors had a strong adverse impact on the overall accuracy of the real GNP forecasts, as indicated by the following tabulation:

	No.	MAE, % Points	% of Total Absolute Error
Underestimation errors	33	1.12	46.8
Overestimation errors	21	.92	24.4
Turning-point errors	8	2.85	28.8

This evidence contradicts the argument that turning-point errors matter little because they are few and far between (see, e.g., Samuelson 1976). But the argument goes further to say that such few large errors are the necessary (and small) price to pay for the avoidance of many large errors "between turning points" by means of optimal estimation procedures such as least squares. However, it is not clear that these procedures imply more than that the variance of the predicted changes must be less than that of the actual changes (and progressively declining as the forecast span is lengthened). It has never been demonstrated that a trade-off between errors at major turning points and other errors exists, and it would seem a counsel of despair for the forecasters to accept it as inevitable. Prediction of cyclical turns in such series as real GNP, though certainly difficult, is not necessarily impossible, particularly on an annual basis (note the good record in forecasting troughs). In sum, there are indeed strong reasons for makers and users of economic forecasts to give a great deal of attention to turning-point errors. Actually, most of them realize this, as is shown by the acknowledged need for and practice of analyzing such errors (Hickman 1972; Adams and Duggal 1974; Fromm and Klein 1976). However, there is certainly much need for improvement here, and room for some new initiatives (e.g., on how to use current signals from leading indicators, see Vaccara and Zarnowitz 1978).[4]

The worst single year for the predictions covered in table 14.2 was 1974, on the eve of which forecasters across the field missed the onset of a serious recession. This and the smaller turning-point errors for 1970 are the main reasons for the rise in the average errors of these forecasts in 1969–76 compared with the earlier years. But the rise in the absolute errors was not large, and it is well to recall that at the same time there was some improvement in accuracy as measured by the criteria of comparisons with extrapolations and correlations of predicted with actual changes (compare lines 4–6 and 7–12, cols. 3–4 and 6–7, in table 14.2). Limited evidence from one longer series of forecasts suggests that real GNP was predicted with similar average errors in the two 8-year periods, 1953–60 and 1969–76, with much smaller errors in the relatively quiet years 1961–68.

Although the forecasts of real GNP are about as good relative to our simple extrapolative benchmark models as are the forecasts of GNP in current dollars, they are less accurate in terms of comparisons of the errors with the actual percentage changes to be predicted. The point is that the extrapolations perform substantially better for nominal GNP than for real GNP. This can be shown by dividing the error of extrapolation into the size of the actual change, without regard to sign, which gives the following overall ratios for the $X1$ (last change) and $X2$ (average change) models:

GNP—$X1$, 0.44; $X2$, 0.30. Real GNP—$X1$, 0.78; $X2$, 0.68.

These results accord with expectations, since the growth rates in constant-dollar GNP varied considerably more than those in current-dollar GNP. The ratios of forecast error to extrapolation error average about 0.4 when $X1$ is the standard, 0.5–0.6 when $X2$ is, and the results are much the same for either variable.

Table 14.3 surveys the performance of forecasts of percentage changes in the price level (IPD) that match the real GNP predictions covered in table 14.2. On the average, the predicted inflation rates fall short of the actual ones by fractions of 1 percentage point (lines 1–3). The 1959–67 forecast sets are less accurate than simple last-change extrapolations (line 4), and the other sets

4. Eckstein (1978, p. 321) expresses doubt on this score, saying that "for example, leading indicators began to fall in August 1974. . . ." But this applies only to the indicators in current dollars; the present leading index based largely on series in constant dollars or physical units turned down in June 1973, 5 months before the date of the NBER reference peak, and kept falling. Similarly, nominal GNP declined only slightly for a single quarter in 1975, quarter 1 (hereafter indicated as 1975:1), while real GNP declined for six quarters between 1973:4 and 1975:1.

Table 14.3 **Summary Measures of Error for Annual Predictions of Percentage Changes in the Price Level, 1959–76**

Line	Period and No. of Years Covered	Selected Private Forecasts, Mean[a] (1)	ASA-NBER Survey, Median Forecast (2)	*Economic Report of the President*[b] (3)	Michigan Model (4)	Wharton Model (5)	Extrapolations: Last Change (6)	Extrapolations: Average Change (7)	Actual: Preliminary (8)	Actual: Revised (9)
	Mean Absolute Percentage Change: Predicted and Actual									
1	1959–67 (9)	1.5	...	...	1.9	...	...	...	1.9(1.4)[c]	2.0
2	1962–76 (15)	...	...	3.7	3.8	...	...	...	4.2	4.5
3	1969–76 (8)	...	4.9	5.3	5.0	5.3	...	...	5.9	6.2
	Mean Absolute Error, in Percentage Points									
4	1959–67 (9)	.6	...	...	.7	...	.3	.7	.3	...
5	1962–76 (15)	...	...	1.0	1.0	...	1.3	1.4	.4	...
6	1969–76 (8)	...	1.3	1.4	1.4	1.4	2.0	2.1	.4	...
	Mean Error, in Percentage Points									
7	1959–67 (9)	.2	...	...	0	...	−.1	.04	−.3	...
8	1962–76 (15)	...	...	−.5	−.5	...	−.2	−1.0	−.3	...
9	1969–76 (8)	...	−.9	−.6	−.9	−.6	−.2	−1.2	−.4	...
	Squared Correlation (r^2) *between Predicted and Actual Change*									
10	1959–67 (9)	.389	...	...	.424	...	.365	.068[d]	...	...
11	1962–76 (15)	...	...	.768	.682	...	.536	.508	...	...
12	1969–76 (8)	...	.526	.581	.454	.604	.166	.059	...	...

Note: For sources and explanations of the data used in cols. 2–9, see nn. *d–k*, respectively, in table 14.1.

[a]Average of end-of-year annual forecasts of the composite price level (a weighted sum of forecasts of CPI and WPI).

[b]See table 14.2, n. *b*.

[c]The figure in parentheses is based on weighted averages of data for CPI and WPI. This series of actuals for the composite price level is comparable to the forecasts used in col. 1 only (cf. table 14.2, n. *c*).

[d]r is negative.

outperform the naive models by relatively small margins, much less than those observed for the GNP series. The naive models work comparatively well here, with errors averaging about three tenths of the actual changes in IPD. Projections of the last change are in this case better than those of the average change (cols. 6–7), which is the reverse of the situation for GNP in both current and constant dollars. The forecasts underestimated strongly (much more than the last-change extrapolations) the average inflation since 1961 (lines 7–9). The predicted and actual percentage changes in the price level are all positively correlated, but the correlations for 1969–76 are generally lower than their counterparts for GNP and, still more so, for real GNP (lines 10–12).

Forecasts of inflation often have much in common with projections of the last observed rate of inflation. To illustrate, correlations between the errors of these forecasts and the errors of the corresponding extrapolations produce the following r^2 coefficients: Michigan (1959–76), .51; CEA (1962–76), .78; ASA-NBER (1969–76), .95; Wharton (1969–76), .80. For growth rates in real GNP, the correlations between forecast errors and extrapolation errors are also positive but throughout lower, in most cases much lower. These results are not surprising, and they have a positive aspect inasmuch as forecasts should be closer to extrapolations of a given type in those cases where such extrapolations are more effective (for an elaboration, see Mincer and Zarnowitz 1969). However, our comparisons are constrained to naive models which presumably do not represent high standards for economic forecasting. In particular, price-level forecasts that are highly correlated with last-change extrapolations must share the property of the latter to lag a year behind the actual rates of inflation. Indeed, the correlations between the predicted changes and the previous year's actual changes are all positive and high: the r^2 coefficients for the four sets of IPD forecasts listed earlier in this paragraph are .76, .87, .81, and .72, respectively. These are far higher than the r^2 between successive actual changes, which range from .17 to .54 depending on the period (col. 6).

The annual percentage changes in real GNP (RGNP) are inversely related to those in IPD and positively related to those in current-dollar GNP, while the last two variables do not show a strong or stable association. The relationships between the predicted changes generally parallel the actual ones. This is illustrated by the r^2 coefficients tabulated below.

	1962–76			1969–76		
	Actual	Michigan	CEA	Actual	Michigan	CEA
RGNP-IPD	.567(−)	.328(−)	.528(−)	.646(−)	.472(−)	.651(−)
RGNP-GNP	.297	.210	.222	.644	.464	.491
IPD-GNP	.020	.217	.068	.085(−)	.004	.022(−)

The errors of the forecasts are similarly interrelated. Table 14.4 demonstrates a pervasive pattern of negative correlation between errors in forecast-

Table 14.4 Correlations between Errors of Forecasts of Percentage Changes in Nominal GNP, Real GNP, and IPD, 1962–76

		Squared Correlation (r^2) between Forecasts Errors		
Line	Source of Forecast	For RGNP and IPD (1)	For RGNP and GNP (2)	For IPD and GNP (3)
1962–76 (15 years):				
1	*Economic Report* (CEA)	.297(−)	.359	.114
2	Michigan model	.494(−)	.429	.006
1969–76 (8 years):				
3	*Economic Report* (CEA)	.677(−)	.004	.259
4	Michigan model	.684(−)	.209	.014
5	Wharton model	.340(−)	.036	.466
6	ASA-NBER survey, median	.525(−)	.013	.351

Note: The sign (−) following the r^2 coefficient indicates that r is negative.

ing real growth and inflation (col. 1). The tendency for these errors to be offsetting, which benefits the forecasts of GNP in current dollars, is most strongly in evidence for the more recent years. When forecasters overestimated real growth, or missed a downturn and projected continued growth instead, they typically also underestimated inflation, as in 1969–71 and 1973–74. Underprediction of real growth occurred in 1972 and 1975–76 in combination with overprediction of inflation.

These observations, which have some precedents (Moore 1969a, 1977a, Zarnowitz 1969b), are consistent with a view of the world in which nominal GNP changes are predicted directly and relatively well, but their division into real and price changes continues to pose great problems. Many forecasters may agree with that view in general terms, and some subscribe to models consistent with it (a specific example might be the St. Louis model, in which the dollar change in total GNP expenditure is determined mainly by the dollar change in a measure of money stock). However, most macroeconometric models, including the two sets covered here, have separate aggregate-real-demand, output, and price-level equations, and it is not at all clear why they should predict GNP better in current than in constant dollars. In fact, some studies of the recent performance of quarterly models arrive at the opposite conclusion, namely, that the results for real GNP are better than those for nominal GNP because of deficient price forecasts (Duggal et al. 1974; Eckstein, Green, and Sinai 1974). The available evidence seems too limited and too mixed to permit any conclusive generalizations on this point. But it is interesting to observe that the importance of output errors versus price errors may vary with changes in the relative roles of real versus nominal factors and disturbances: in the 1970s the errors of the GNP forecasts were for the most part better correlated with the IPD errors than with the RGNP errors, whereas in the 1960s the contrary situation obtained (table 14.4, cols. 2–3).

14.4 Quarterly Multiperiod Forecasts, 1970–75: An Overall Appraisal

Here the forecasts and actual data refer to overlapping sequences of quarters, not simply to a series of successive years. Our materials cover 22 quarters from 1970:3 through 1975:4, a period for which forecasts from several new sources are available. First estimates for the preceding year, taken from the data prior to the 1976 benchmark revision of the national income accounts, serve as comparable realizations. Comparisons with the revised data will require a fully integrated treatment of forecast errors and measurement errors. In short, we now deal with much more complex forecasts but basically treat the predictions and actual data in the same straightforward fashion as before.

Four of the sources use formal macroeconometric models combined with judgmental adjustments and (often several alternative) assumptions about exogenous factors. These include, in alphabetical order: Bureau of Economic Analysis of the U.S. Department of Commerce (BEA); Chase Econometric Associates, Inc. (Chase); Data Resources, Inc. (DRI); and Wharton Econometric Forecasting Associates, Inc. (Wharton). Each of these forecasters predicts more than 100 variables each quarter, for sequences of up to 8 quarters ahead, so each can be said to produce regular and relatively detailed forecasts of the short-term course of the U.S. economy. However, the models used differ in several important respects, such as size and selection of the exogenous variables. Thus, the BEA model is currently viewed as of "medium" scale (it has 98 equations, of which 58 are stochastic); Chase is "large" (it has 150 equations, of which 125 are stochastic); so is Wharton (which has 201 equations, of which 67 are stochastic); and DRI is "very large" indeed (with 718 equations, 379 of which are behavioral).[5]

Forecasts by a group associated with the General Electric Company (GE) use individual econometric relationships in an iterative process involving a large amount of judgmental prediction; they are not based on a formal model of simultaneous equations. This system also is large enough to produce forecasts of over 100 variables. In addition, the ASA-NBER median forecasts are included as a benchmark set of composite predictions. The surveys attract a representative cross section of professional forecasters, mainly business but also academic and government economists. The number of participants varies, averaging about 50 per survey. Their methods and techniques differ, but informal use of models and judgment prevails. The averages of the individual

5. See Intriligator 1978, table 12.12, pp. 454–56, for the numbers of equations and other summary characteristics of the models. Fromm and Klein (1976, table 1, p. 2) define the scale of a model, based on the number of equations, as follows: very small = 9 or less, small = 10–49, medium = 50–119, large = 120–99, very large = 200 or more.

predictions from the surveys trace well the trends in what is sometimes called the "standard" or "consensus" forecast.

As shown in table 14.5, lines 1–6, the MAEs of GNP forecasts are close to 1 percentage point (like the annual forecasts, see table 14.1) for 2 quarters ahead and about half of that or less for 1 quarter ahead. Over longer spans, the MAEs rise more or less steadily by increments varying generally from 0.2 to 0.5 of 1 percentage point for each additional quarter. The MAEs approach 2 percentage points for 4-quarter spans and exceed it for 5-quarter spans; similarly, they approach 3 percentage points for 7-quarter spans and exceed it for 8-quarter spans. Consistent with earlier findings and interpretations for various types of multiperiod forecasts (see, e.g., Zarnowitz 1967, pp. 60–72), the MAEs increase somewhat less than in proportion to the extension of the span. The errors in forecasts of percentage changes expressed on a per-unit-of-time basis (roughly, MAE divided by the length of the effective span) are substantially stable, neither rising nor declining systematically as the forecast reaches further into the future. The same applies to the errors of the implicit predictions of changes during the successive single quarters covered; it is the cumulation of these intraforecast ("marginal") change errors that technically accounts for the tendency of errors in the total predicted changes to grow with the span.[6]

Where both forecasts and realizations refer to increases (as they do most of the time by far in the case of GNP), errors of positive sign denote overestimation and errors of negative sign denote underestimation of actual change. For the short spans of 1–4 quarters, the mean errors (ME) are in all but one instance positive for BEA, Chase, and DRI, while they are, on the contrary, negative throughout for ASA-NBER, GE, and Wharton (table 14.5, lines 7–12). In absolute terms, these averages are predominantly small, however (less than $|0.2|$ except for the ASA-NBER errors, which are larger). For the longer spans of 5–8 quarters, the MEs are generally negative, much larger absolutely, and rising with the span. As will be shown below, the overall MEs conceal large errors of opposite sign in the forecasts for some of the different economic phases of the period 1970–75.

The r^2 coefficients for the correlations between the predicted and actual changes in GNP exceed .6 or .7 for 1-quarter ahead (like the annual forecasts) and exceed .4 or .5 for 2 quarters ahead (lines 13–18). They are generally much smaller for the longer spans, mostly in the .1–.3 range, in a few cases (interestingly, for the middle, rather than the longest, spans) near zero.

Theil's inequality coefficients are remarkably similar for the different forecast sources and horizons, falling mostly between 0.2 and 0.3 (lines 19–24). This indicates that these forecasts are all much better than a naive model ex-

6. Note that fewer observations are available for the longer spans (table 14.5, n. *b*). This reduces the comparability of the measures reported for the different spans but does not eliminate it entirely.

Table 14.5 **Summary Measures of Error for Quarterly Multiperiod Predictions of Percentage Change in GNP, 1970–75**

		Span of Forecast in Quarters[b]							
Line	Forecast Set[a]	One (1)	Two (2)	Three (3)	Four (4)	Five (5)	Six (6)	Seven (7)	Eight (8)
		Mean Absolute Error, in Percentage Points							
1	ASA-NBER	.51	1.03	1.44	1.67	2.10	. . .	. . .	. . .
2	BEA	.46	.84	1.36	1.67	2.27	. . .	. . .	. . .
3	Chase	.42	1.03	1.32	1.68	2.22	2.73	3.19	3.49
4	DRI	.53	1.04	1.43	1.94	2.43	2.69	2.95	2.80
5	GE	.42	.95	1.34	1.71	2.19	2.59	2.88	3.25
6	Wharton	.40	.98	1.60	1.68	1.92	2.35	2.83	3.07
		Mean Error, in Percentage Points[d]							
7	ASA-NBER	.06	−.27	−.38	−.39	−.91	. . .	. . .	. . .
8	BEA	.05	.11	.10	.11	−.22	. . .	. . .	. . .
9	Chase	.01	.04	.02	.08	−.14	−.66	−1.48	−2.34
10	DRI	−.01	.11	.05	.11	.01	−.42	−1.12	−1.69
11	GE	−.14	−.15	−.30	−.15	−.15	−.44	−.95	−1.68
12	Wharton	−.12	−.10	−.19	−.12	−.15	−.69	−1.32	−2.17
		Squared correlation (r^2) between Predicted and Actual Changes							
13	ASA-NBER	.612	.615	.262	.168	.384	. . .	. . .	. . .
14	BEA	.704	.597	.201	.032	.007	. . .	. . .	. . .
15	Chase	.752	.451	.107	.058	.127	.134	.179	.293
16	DRI	.632	.469	.069	.000[e]	.008	.102	.249	.600
17	GE	.753	.577	.284	.159	.132	.180	.227	.225
18	Wharton	.773	.453	.099	.148	.252	.349	.310	.440
		Theil's Inequality Coefficient (U)							
19	ASA-NBER	287	.270	.266	.242	.210	. . .	. . .	. . .
20	BEA	.250	.256	.271	.266	.251	. . .	. . .	. . .
21	Chase	.241	.287	.294	.268	.236	.218	.198	.189
22	DRI	.284	.292	.299	.295	.272	.218	.182	.148
23	GE	.245	.260	.260	.243	.233	.204	.181	.172
24	Wharton	.223	.291	.306	.249	.215	.186	.183	.171

[a]See text on the sources and abbreviations used.

[b]Number of forecasts covered per set: ASA-NBER, BEA; 22, 21, 20, 19, and 17 for spans 1–5, respectively; Chase, DRI, GE, Wharton; 22, 21, 20, 19, 18, 17, and 16 for spans 1–7, respectively; 15 (Chase, Wharton), 14 (DRI), and 12 (GE) for span 8.

[c]MAE $= 1/n\Sigma|e_t|$, where $e_t = P_t - A_t$, P_t = predicted percentage change, A_t = actual percentage change, and n = number of forecasts covered (for the given set, variable, period, and span).

[d]ME $= 1/n\Sigma e_t$ (see n. *c*).

[e]*r* is negative.

[f]$U = \sqrt{\Sigma e_t^2/\Sigma A_t^2}$ (see n. *c*).

trapolating the last-recorded percentage change (for which $U = 1$). That model, it should be noted, is but a minimal standard for economic forecasts. It is worth noting that the U coefficients do not increase with the forecast span; in fact, they decline slightly below .2 for the longest spans.

The next two tables have the same format as table 14.5, which facilitates presentation and comparisons of these measures. Real GNP have MAEs (in percentage points) rising from .5–.6 for 1 quarter ahead to 5–6 for 8 quarters ahead, that is, somewhat more than in proportion to the measured span (table 14.6, lines 1–6). The errors for the two shortest spans are not much larger than those for GNP in current dollars, but the errors for the longest spans are 50%–100% larger. The unusually rapid buildup of the MAEs can be traced in large part to turning-point errors. In quarterly multiperiod forecasting, turning points are more frequent and more difficult to predict than in annual forecasting, but the errors associated with them matter much more yet. Here, missing a turn often means that a whole chain of predictions for the subsequent observations is badly off.

The MEs of these forecasts are all positive, which is largely due to the effects of missing or underestimating the declines in real GNP during the recession (table 14.6, lines 7–12). The MEs also cumulate continuously and rapidly here, quite unlike those for the nominal GNP forecasts.

Relative to the size of the actual percentage changes, the real GNP errors are on the average much larger than the current-dollar GNP errors. Note that the inequality coefficients in table 14.6, lines 19–24, rise from .4 or more for the shortest to .7 or more for the longest forecasts. (In contrast, let us recall, the corresponding statistics in table 14.5 vary in the vicinity of .2 for all spans.)

In terms of correlations with the actual changes, on the other hand, the real GNP forecasts look favorable. The r^2 coefficients are throughout rather surprisingly high in table 14.6, averaging over .8 for 1 quarter ahead, over .7 for 2, 5, and 6 quarters ahead, and over .6 for the other spans (lines 13–18). They are much higher than the r^2 coefficients for GNP in table 14.5, particularly for the longer forecasts.

The MAEs of forecasts of inflation in terms of the GNP IPD are once more of the order of .5 or less 1 quarter ahead, approximately 1 percentage point 2 quarters ahead (table 14.7, lines 1–6). However, they cumulate very rapidly, especially in the range of the longer forecasts. The figures for the 4-quarter-ahead predictions are here on the average about 5 times as large and the figures for 8 quarters ahead are more than 13 times as large as those for the 1-quarter span.

This exceptionally strong, much more than proportional buildup of errors reflects a progression of underestimates of inflation rates. The MEs in table 14.7 are all negative (lines 7–12) and average about $-.2$, -1.5, and -5.5 percentage points for the spans of 1, 4, and 8 quarters, respectively.

The r^2 coefficients for the IPD inflation forecasts are generally higher than

Table 14.6 **Summary Measures of Error for Quarterly Multiperiod Predictions of Percentage Change in Real GNP, 1970–75**

		Span of Forecast in Quarters[b]							
Line	Forecast Set[a]	One (1)	Two (2)	Three (3)	Four (4)	Five (5)	Six (6)	Seven (7)	Eight (8)
		Mean Absolute Error, in Percentage Points[c]							
1	ASA-NBER	.56	1.24	1.89	2.43	3.09	...	...	...
2	BEA	.56	1.37	2.14	2.80	3.17	...	...	...
3	Chase	.51	1.11	1.81	2.46	3.29	4.19	4.95	5.31
4	DRI	.61	1.37	2.08	2.75	3.52	4.15	4.78	5.58
5	GE	.50	1.20	1.75	2.15	2.80	3.80	4.76	5.15
6	Wharton	.45	1.18	1.73	2.00	2.48	3.18	4.24	4.92
		Mean Error, in Percentage Points[d]							
7	ASA-NBER	.10	.32	.76	1.30	1.71	...	...	...
8	BEA	.32	.68	1.12	1.71	1.83	...	...	...
9	Chase	.17	.51	.92	1.46	1.98	2.38	2.63	2.82
10	DRI	.26	.77	1.20	1.82	2.59	3.16	3.66	4.72
11	GE	.00	.22	.36	.95	1.53	2.09	2.46	2.58
12	Wharton	.02	.35	.72	1.22	1.71	2.17	2.69	2.90
		Squared Correlation (r^2) between Predicted and Actual Change							
13	ASA-NBER	.846	.789	.698	.684	.746	...	...	...
14	BEA	.829	.754	.642	.512	.508	...	...	...
15	Chase	.839	.817	.727	.703	.733	.710	.604	.596
16	DRI	.793	.745	.598	.584	.785	.827	.741	.638
17	GE	.808	.741	.677	.661	.772	.764	.661	.662
18	Wharton	.820	.672	.651	.745	.816	.781	.730	.580
		Theil's Inequality Coefficient (U)[e]							
19	ASA-NBER	.461	.532	.618	.663	.660	...	...	...
20	BEA	.476	.585	.675	.756	.701	...	...	...
21	Chase	.433	.502	.607	.673	.711	.741	.758	.741
22	DRI	.504	.622	.721	.769	.781	.774	.774	.836
23	GE	.427	.498	.548	.606	.627	.676	.714	.694
24	Wharton	.398	.543	.582	.563	.555	.584	.632	.662

[a]See text on the sources and abbreviations used.

[b]Number of forecasts covered per set: ASA-NBER, BEA: 22, 21, 20, 19, and 17 for span 1–5, respectively; Chase, DRI, GE, Wharton: 22, 21, 20, 19, 18, 17 and 16 for spans 1–7, respectively; 15 (Chase, Wharton), 14 (DRI), and 12 (GE) for span 8.

[c]MAE $= 1/n\Sigma|e_t|$, where $e_t = P_t - A_t$, P_t = predicted percentage change. A_t = actual percentage change, and n = number of forecasts covered (for the given set, variable, period, and span).

[d]ME $= 1/n\Sigma e_t$, (see n. *c*).

[e]$U = \sqrt{\Sigma e_t^2/\Sigma A_t^2}$ (see n. *c*).

those for the forecasts of percentage change in nominal GNP (except for a few short predictions) but throughout lower than the corresponding statistics for the real-growth forecasts. They range from .23 to .71 and tend to decrease as the spans lengthen (table 14.7, lines 13–18).

The inequality coefficients U average .31 for the shortest and .48 for the

Table 14.7 **Summary Measures of Error for Quarterly Multiperiod Predictions of Percentage Change in the Price Level, 1970–75**

		Span of Forecast, in Quarters[b]							
Line	Forecast Set[a]	One (1)	Two (2)	Three (3)	Four (4)	Five (5)	Six (6)	Seven (7)	Eight (8)
		Mean Absolute Error, in Percentage Points[c]							
1	ASA-NBER	.45	1.02	1.64	2.26	3.14	. . .	. . .	. . .
2	BEA	.44	1.10	1.86	2.55	2.96	. . .	. . .	. . .
3	Chase	.39	1.02	1.64	2.29	2.98	3.87	4.88	5.69
4	DRI	.54	1.11	1.69	2.37	3.05	4.04	5.17	6.78
5	GE	.39	.90	1.49	1.96	2.37	3.06	4.08	4.79
6	Wharton	.35	.79	1.38	1.95	2.76	3.57	4.54	5.53
		Mean Error, in Percentage Points[d]							
7	ASA-NBER	−.26	−.64	−1.20	−1.82	−2.77	. . .	. . .	. . .
8	BEA	−.24	−.57	−1.07	−1.70	−2.18	. . .	. . .	. . .
9	Chase	−.15	−.49	−.96	−1.50	−2.33	−3.31	−4.42	−5.48
10	DRI	−.27	−.70	−1.22	−1.85	−2.82	−3.88	−5.12	−6.78
11	GE	−.12	−.36	−.70	−1.20	−1.85	−2.78	−3.76	−4.57
12	Wharton	−.14	−.46	−.96	−1.44	−2.02	−3.08	−4.29	−5.34
		Squared Correlation (r^2) between Predicted and Actual Change							
13	ASA-NBER	.651	.587	.474	.410	.385	. . .	. . .	. . .
14	BEA	.574	.388	.309	.217	.210	. . .	. . .	. . .
15	Chase	.600	.440	.394	.287	.246	.233	.320	.381
16	DRI	.478	.426	.412	.346	.401	.398	.384	.371
17	GE	.657	.633	.508	.440	.438	.457	.524	.676
18	Wharton	.714	.711	.652	.566	.502	.495	.494	.396
		Theil's Inequality Coefficient U)[e]							
19	ASA-NBER	.325	.341	.382	.408	.458	. . .	. . .	. . .
20	BEA	.338	.382	.410	.445	.446	. . .	. . .	. . .
21	Chase	.311	.358	.377	.410	.438	.462	.480	.496
22	DRI	.375	.388	.397	.422	.444	.475	.508	.540
23	GE	.284	.286	.326	.354	.369	.395	.416	.410
24	Wharton	.265	.270	.307	.339	.365	.399	.436	.465

[a]See text on the sources and abbreviations used.

[b]Number of forecasts covered per set: ASA-NBER, BEA: 22, 21, 20, 19, and 17 for spans 1–5, respectively; Chase, DRI, GE, Wharton: 22, 21, 20, 19, 18, 17, and 16 for spans 1–7, respectively; 15 (Chase, Wharton), 14 (DRI), and 12 (GE) for span 8.

[c]MAE $= 1/n\Sigma|e_t|$, where $e_t = P_t - A_t$, P_t = predicted percentage change, A_t = actual percentage change, and n = number of forecasts covered (for the given set, variable, period, and span).

[d]ME $= 1/n\Sigma e_t$ (see n. *c*).

[e]$U = \sqrt{\Sigma e_t^2/\Sigma A_t^2}$ (see n. *c*).

longest inflation forecasts, displaying a systematic but gradual increase with the span (lines 19–24). They are higher than their counterparts for the current-dollar GNP forecasts but lower than those for the real GNP forecasts.

A critical point revealed by the comparison of our tabulations is that the quantity and price ingredients of the GNP forecasts show a pattern of offsetting errors in the quarterly as well as annual data. The MEs of real-growth predictions are all positive; those of inflation predictions, all negative. Matched by source and span, these statistics have similar absolute values for most of the shorter forecasts. As a result, the MEs of the current-dollar GNP forecasts for spans 1–4 vary in sign but are generally small. For the longer spans 5–8, however, the negative MEs of the inflation forecasts outweigh the positive MEs of the real-growth forecasts, often by large margins. Consequently, the MEs of the longer forecasts of percentage changes in GNP are almost without exception negative (denoting underestimation) and for the most part relatively large.

Forecasts from the different sources are comparable only in a limited and qualified sense. Not only do the models vary in scale and choice of exogenous variables (and practically nothing can be done to allow for these differences), but there are also some systematic discrepancies in the timing of the predictions. The GE forecasts are typically made early in the quarter.[7] Late forecasts are often somewhat more accurate than those released earlier, since they enjoy advantages of additional and more current information.[8] Some of the sources produce forecasts 2 or even 3 times a quarter, but so far we have been able to obtain only one set of predictions for each of them; with more information, it will be possible to increase the number of sets which are more nearly comparable in terms of the release dates.[9]

Despite the many difficulties that beset them, comparisons between forecasts are needed and will inevitably be made. Some summary statements on the apparent relative accuracy of the forecasts are therefore in order here, but they must be interpreted with great caution and cannot support any strong conclusions. According to the MAEs, Wharton and GE forecasts of both real growth and inflation rank high (i.e., they have relatively low MAEs), but as already noted, these sets, particularly GE, have the advantage of late timing. Among the four "early-quarter" forecasters, Chase and ASA have better records than BEA and DRI for most of the 1–5-quarter forecasts of percentage changes in real GNP and IPD, but BEA ranks overall higher than the others for the corresponding nominal GNP predictions. Of the four, only Chase and DRI predict over longer (6–8-quarter) spans, and here DRI has somewhat lower errors for the growth rates in nominal and real GNP, and Chase has somewhat lower errors for inflation.

The differences between the MAEs across sources are mostly small; their precise significance is unknown but probably in many cases low. Other mea-

7. For the release dates, see McNees 1975, p. 39; 1976, p. 41.

8. For evidence, see Zarnowitz 1967, pp. 126–30 and McNees 1975, pp. 22–30; 1976, pp. 31 ff.

9. Three of the sources predict more than once each quarter, and the calendar of the releases suggests the following classification for them (EQ, MQ, and LQ denote early-quarter, mid-quarter, and late-quarter forecasts, respectively):

Chase: EQ, LQ   DRI: EQ, MQ, LQ   Wharton: EQ, MQ sures of forecasting performance such as the correlations between predicted and actual changes often yield different rankings of the forecast sets. I conclude that the surveyed measures do not show any of the forecasters to be consistently and generally superior to others. This result is consistent with other evaluations which find that the rankings of forecasters vary depending on the variables, periods, and spans covered as well as on the criteria and measurements applied.[10]

14.5 Quarterly Multiperiod Forecasts: An Analysis by Subperiods

The period 1970:3–1975:4, although short, was unusually varied and marked by major disturbances and drastic changes in the economy's course. It is useful to divide it into the following parts, as suggested by the contemporary business cycle and inflationary developments: (1) *1970:3–1973:1:* End of the mild 1970 recession followed by an expansion that accelerated in 1972, with relatively stable inflation. (2) *1973:1–1973:4:* Slower real growth and a sharp inflation speedup (materials shortages, runups in commodity prices, oil embargo). (3) *1973:4–1975:1:* Recession, severe in its last 2 quarters, accompanied first by a further rise and then by a downturn in the rate of inflation. (4) *1975:1–1975:4:* Sharp upturn and the initial recovery phase, with a further decline in inflation.

One question is whether forecasts that originated in these four subperiods show significantly different characteristics and performance. The other is whether forecasts for these subperiods (i.e., those that aimed at the corresponding groups of target quarters) are so differentiated. It turns out that the answers to both questions are definitely yes.

To illustrate the first point, the expansion phase 1 produced forecasts that underestimated growth in dollar GNP mainly because they underestimated inflation. The percentage changes in real GNP were partly underpredicted, partly (in some longer forecasts) overpredicted, but whether negative or positive the MEs of these forecasts were small. In general, the record of the forecasts that were made during period 1 was good in terms of both the ME and the MAE figures, even for the long spans. In contrast, the slowdown phase 2 produced real-growth predictions with very large positive MEs and inflation forecasts with very large negative MEs (underestimation errors). These errors balanced each other so that the MEs for the nominal GNP predictions were moderate (and mostly negative, except for the longest forecasts). The reces-

10. The findings of this study should and will be carefully compared with those of other appraisals of the same forecasts, but the task is still to be completed. Differences between the error measures used are difficult to allow for and they becloud the results of such comparisons. Nonetheless, it is interesting to observe that the 1975 and 1976 studies by McNees report rankings of forecasting performance for GNP, real GNP, and IPD that resemble rather well the results of the present evaluation. (McNees also shows that the forecasters often score quite differently for other variables.)

sion phase 3 gave rise to even larger positive MEs in the real-growth forecasts as the declines were repeatedly missed and, when finally recognized, underestimated. These errors were larger absolutely than the negative errors on the price side, which reflected a continuing underestimation of inflation, so that the predictions of the growth rates in nominal GNP had consistently positive MEs in subperiod 3.

The above summary is based on charts (not reproduced here) which show the average errors (MAE and ME) by span and by subperiod in which the forecasts originated.[11] These charts look very similar for the different models. They show in each case the same striking differences between the forecasts made in subperiods 1, 2, and 3. The suggested inference is that concurrent predictions from different sources and models have common patterns such that their errors depend strongly and similarly on the characteristics of the time of their origin.

In a second exercise, the forecasts were assigned to the four subperiods according to their target quarters, not their base quarters, as illustrated in figures 14.1–14.3. Here the samples are partitioned differently; hence the resulting patterns diverge from those obtained on the first plan, but the conclusion is analogous. The type and size of forecast errors depend critically on the economic properties of the target periods vis-à-vis those of the periods of origin. Forecasters perform best when the two periods are alike, belonging to the same already recognized phase—for example, a continuing expansion as in 1971–72 (most of subperiod 1). They perform worst when the target falls into a new phase, particularly when the latter departs sharply from the currently established pattern (forecasts made in subperiods 2 and 3 and those for subperiods 3 and 4 provide many examples, particularly in the long-span categories). Such period characteristics are much more important determinants of forecast errors than are any differences among the forecasters.

Figure 14.1 shows that all six forecasters persistently underestimated the percentage changes in GNP for subperiods 1 and, much more strongly, 2. All of them overestimated GNP changes in their short forecasts for subperiod 3 (from 1 to 3 or 4 quarters ahead) and underestimated changes in the longer forecasts for the same phase. Overestimates prevailed in all sets for the last phase covered, subperiod 4. As would be expected, MAEs typically increase with the span in any phase, and so do MEs when taken without regard to sign. In subperiod 4, however, the average errors behave in an unusual fashion, first increasing and then decreasing with the lengthening span (the rises refer to spans 1–5; the declines to spans 6–8). This is due to offsets between the real-growth forecasts with positive MEs and the inflation forecasts with negative MEs (see figs. 14.2 and 14.3).

Overall, figure 14.1 demonstrates for the GNP forecasts that the average

11. No averages for phase 4 are used on this basis, since they contain too few observations in the truncated sample.

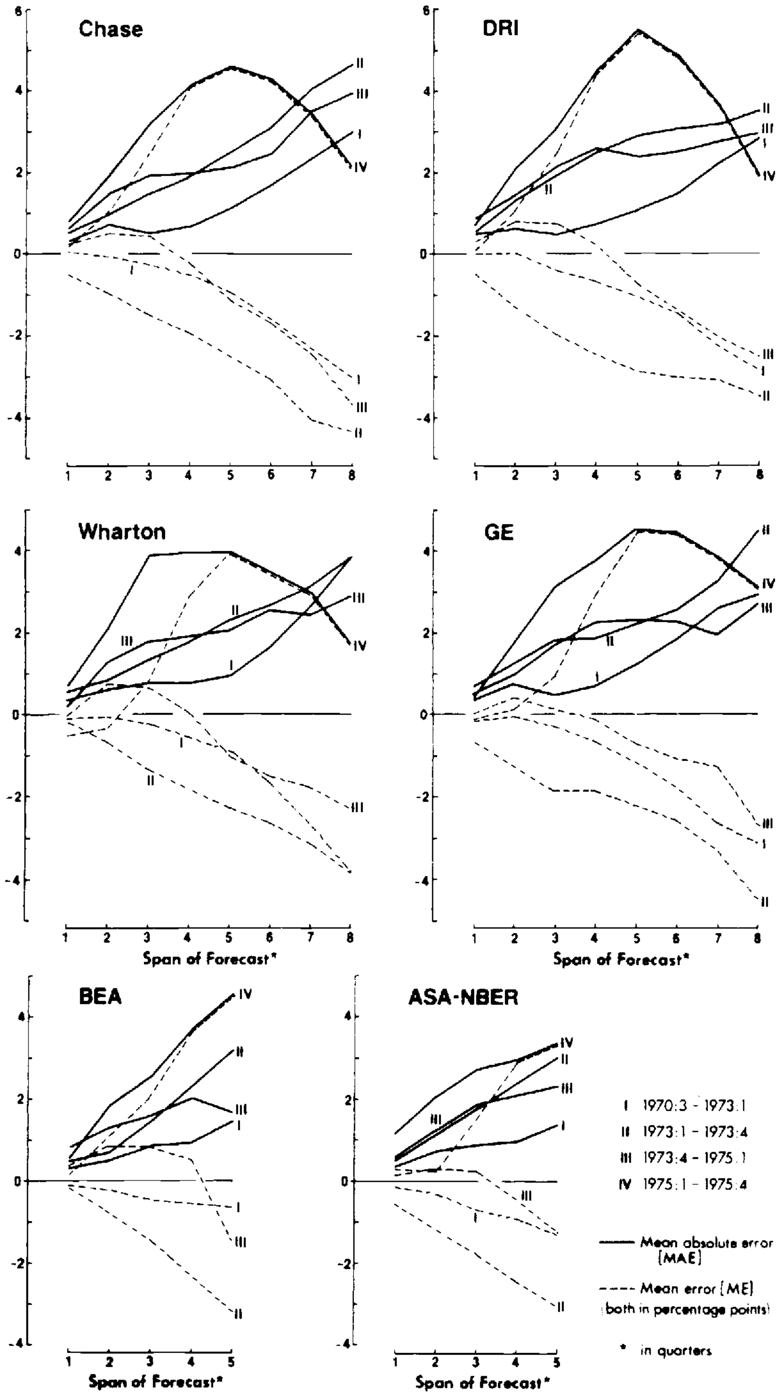


Fig. 14.1 Quarterly multiperiod forecasts of percentage changes in GNP, average errors by subperiod and span, two models, 1970–75

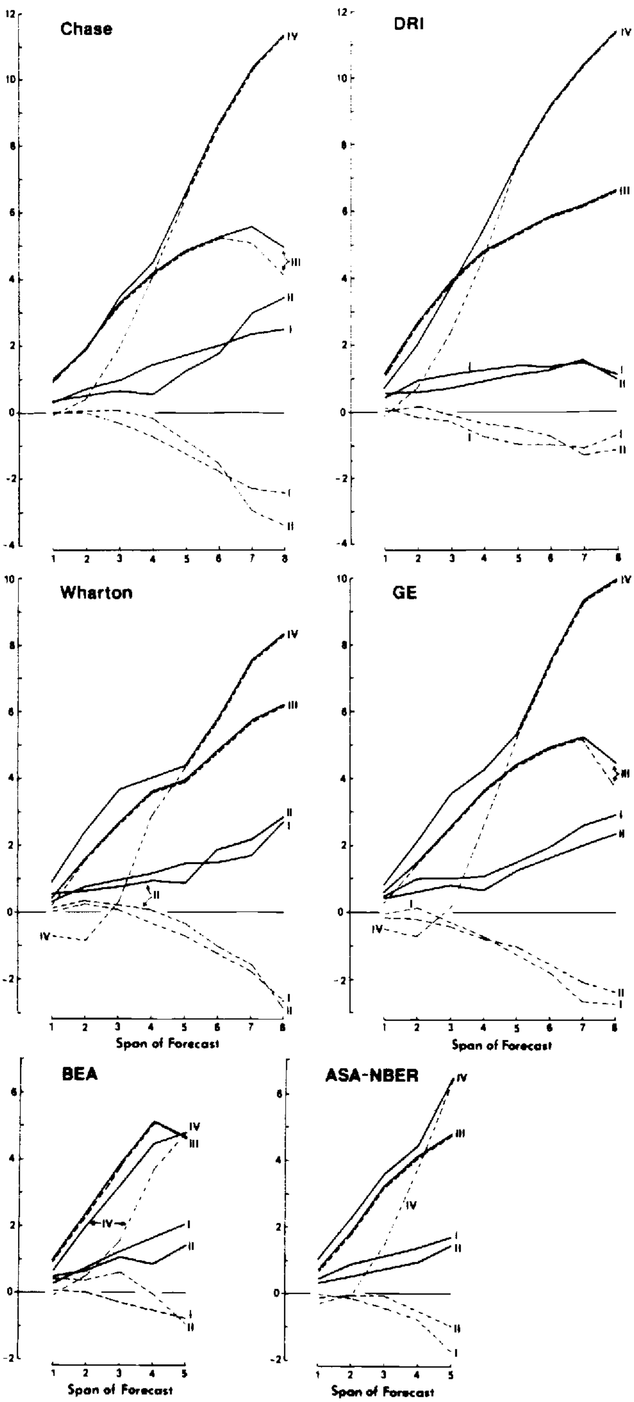
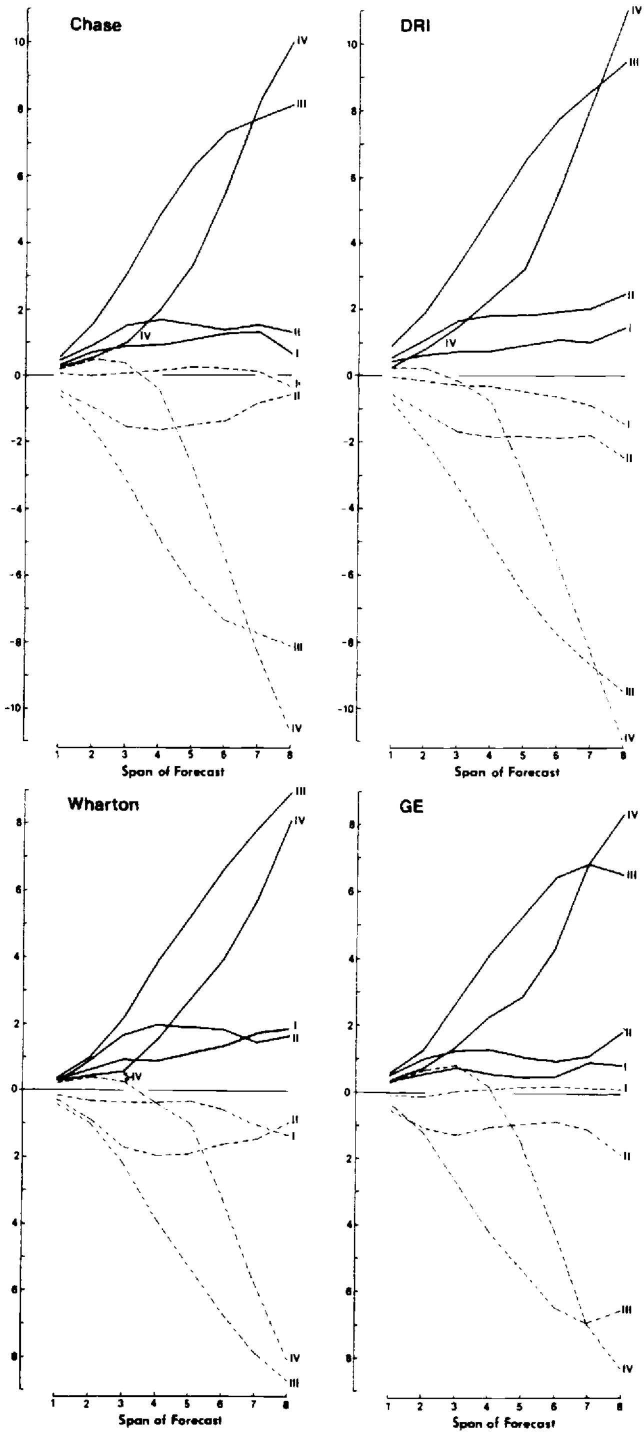


Fig. 14.2 Quarterly multiperiod forecasts of percentage changes in real GNP, average errors by subperiod and span, two models, 1970–75

Note: For key, see fig. 14.1.


Chase
DRI
Wharton
GE
Span of Forecast
Span of Forecast
Span of Forecast
Span of Forecast
IV
III
II
I

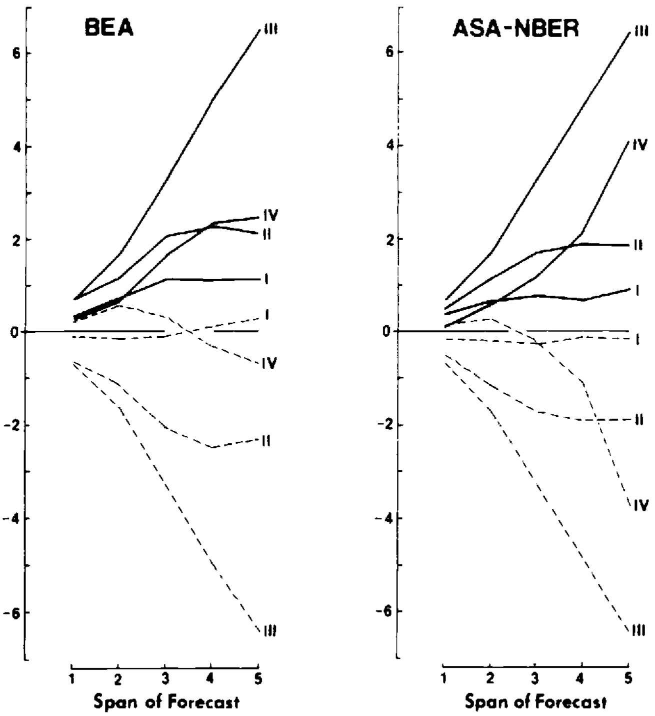


Fig. 14.3 Quarterly multiperiod forecasts of percentage changes in the IPD, average errors by subperiod and span, two models, 1970–75
Note: For key, see fig. 14.1.

error patterns by subperiod and span are substantially similar across the different sources covered. As will be seen below, much the same statement applies to the MAE and ME patterns for the real GNP and IPD forecasts.

Figure 14.2, which covers the real-growth forecasts, shows underestimates dominating the relatively moderate errors for the expansion and slowdown phases 1 and 2. The errors for the recession phase 3 and, particularly, for the 1975 recovery phase 4 are much larger, and they cumulate far more rapidly over the longer spans. The MEs for the two latter subperiods are positive and, for the long forecasts, as large as the MAEs or nearly so. Here the huge errors derive mainly from the forecasters' failure to predict the declines in real GNP and, to a lesser extent, from underestimation of the declines that were predicted.[12] The buildup of errors tapers off at the long end of the forecast range

12. The change errors, $P_t - A_t$ (see table 14.5 for the definitions), are positive where $P_t > 0$ and $A_t > 0$ and also where $P_t < 0$ and $A_t < 0$ but $|P_t| < |A_t|$. These cases dominate in fig. 14.2 the results for both the recession phase 3 and the recovery phase 4. Although real GNP reached a trough in 1975:1 and increased thereafter, the actual changes over longer spans ending in 1975 are negative: that is, real GNP was lower during period 4 than in 1973 (period 2) and during most of 1974 (period 3).

in subperiod 3, but it is strikingly fast throughout in subperiod 4, where the MAEs and MEs increase tenfold or more between the 1-quarter span and the 8-quarter span.

Figure 14.3 shows that the inflation forecasts had relatively small errors in subperiod 1 and considerably larger and on the average all negative errors in subperiod 2; in neither phase did the errors increase strongly with the span. In contrast, the recession period 3 witnessed very large inflation errors, dominated throughout by underestimates and cumulating rapidly. The short forecasts for the initial recovery phase 4 had small errors, mostly with positive means, indicating overestimation of inflation rates that just began to decline. The long forecasts for this last subperiod, however, show very large underestimation errors.

A close examination of figures 14.1–14.3 discloses certain appreciable differences between the forecasters' performances with respect to the particular variables, subperiods, and spans covered. For example, DRI has smaller errors than Chase in predicting real growth 5–8 quarters ahead for phases 1 and 2, but larger errors than Chase in the corresponding forecasts for phases 3 and 4; in predicting growth over the short range of 1–4 quarters, Chase is on balance ahead of DRI. The average accuracy record of inflation forecasts favors GE over Wharton for phase 1 and Wharton over GE for phase 4; for 2 and 3, the picture is mixed, with Wharton scoring as well as GE, or a little better, in the short range but worse than GE in the long range. However, all of these are matters of detail, and the differences are typically small. The main lesson from the comparisons of the MAE and ME patterns is that the similarities greatly outweigh the differences between the forecasters' performance records.

14.6 Concluding Observations

The end-of-year forecasts of annual percentage changes in GNP earn good marks for overall accuracy when judged according to realistic, rather than ideal, standards. Moreover, they are found to have improved in the period since the early 1960s compared with the previous years after World War II.

The corresponding forecasts for GNP in constant dollars (real growth) and IPD (inflation) are weaker. The former suffer from large turning-point errors; the latter from large underestimation errors. But the errors in forecasts of real growth are negatively correlated with the errors in forecasts of inflation, which helped to make the nominal GNP predictions more accurate. In recent times, these correlations were connected with the unexpected concurrence of accelerating inflation and slowing, then declining output rates. Optimistically, and probably also from a lingering faith in a simple Phillips trade-off, forecasters kept anticipating less inflation and more growth. But in the late 1950s and early 1960s, it was the relative stability of the price level that caused widespread surprises, and offsetting errors resulted from the opposite combination of overestimates of inflation and underestimates of real growth.

Forecasts of inflation are not much better than projections of the most recently observed inflation rates, and they lag behind the actual rates much like such projections. The deficiency of price-level forecasts, documented in this and other studies, surely impairs the general ability of economists to analyze the prospects for the economy. Improvements will require major advances in our knowledge, presumably through research based on carefully worked out data (abstract speculation abounds, but good information and observation are rare in this area).

The favorable record of annual GNP predictions does not imply that forecasters can perform well the more difficult task of predicting quarterly changes in GNP within the year ahead or even beyond it. Forecasts for the year as a whole can be satisfactory when based on a good record for the first 2 quarters; they tend to be more accurate than forecasts with longer spans.[13] An examination of the recent multiperiod predictions shows that the errors for real GNP and IPD cumulated rapidly beyond the spans of 2–4 quarters. Previous studies have shown the cumulation to be as a rule less than proportional to the increase in the span, but in this period the buildup of errors was much greater than usual. No doubt in less turbulent times the longer forecasts can be considerably more accurate, but this fair-weather argument is not very persuasive or helpful. At the present time, the predictive value of detailed forecasts reaching out further than a few quarters ahead must be rather heavily discounted. Again, what is critical here is theoretical analysis and empirical research that would lead to improvements in our ability to predict broad movements in the price level and business cycle turning points. Despite setbacks, there is still no reason to give up moderate hopes for an ultimate advance on these fronts.

13. Also, errors of predictions for the individual parts of the year at times offset each other to some degree (Zarnowitz 1967; McNees 1973, 1974). These gains from aggregation over time resemble those from aggregation over sectors (GNP is predicted with smaller average errors of relative change than are most of its components; see Zarnowitz 1967, 1972b; Fromm and Klein 1976).