View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Research Papers in Economics

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

Volume Title: Economic Research: Retrospect and Prospect Vol 1: The Business Cycle Today

Volume Author/Editor: Victor Zarnowitz, ed.

Volume Publisher: UMI

Volume ISBN: 0-87014-251-8

Volume URL: http://www.nber.org/books/zarn72-1

Publication Date: 1972

Chapter Title: Forecasting Economic Conditions: The Record And The Prospect

Chapter Author: Victor Zarnowitz

Chapter URL: http://www.nber.org/chapters/c4397

Chapter pages in book: (p. 183 - 240)

Forecasting Economic Conditions: The Record and the Prospect

Victor Zarnowitz
University of Chicago and National Bureau of Economic Research

1. Introduction: Uses, Sources, and Collection of Forecasts

To men who must plan and act——in government, business, even in such private affairs as personal finance——economic forecasts are tools for reducing uncertainty and inputs into the process of making decisions. To persons in positions of authority, they may serve as a means of communicating intentions or influencing opinion. To professional economists, they are important as products of theories, judgments, and estimating procedures, which can be used in testing the underlying hypotheses, models, and methods. Forecasts of business conditions can be sampled to examine their consensus and dispersion at any time and also continuous revisions over time: this may provide the observer of the economic scene with useful information about what the prevailing climate of opinion is and how it is changing. By analyzing the relations between predictions and subsequent realizations, the accuracy of the former, which interests both makers and users of forecasts, can be assessed. By analyzing the relations between current predictions and earlier predictions and actual events, some understanding can be gained of the genesis of those forecasts that are not based on specific, reproducible methods, and this is of particular concern to the student of the formation and economics of expectations.

The recent and current studies of short-term economic forecasting by the National Bureau of Economic Research, which supply much material for this

Note: I wish to thank Messrs. Mervin A. Daub and James C. Ellert and Mrs. Josephine Su for their valuable statistical assistance. I am also grateful to Mrs. Ester Moskowitz, who edited the manuscript.

paper, are directly or indirectly concerned with all of these different uses of forecasts. The reports by Moore and Shiskin, Zarnowitz, Fels and Hinshaw, Mincer, Cole, and others, produced in the first phase of this project and published during 1967–69,[1] focus on the accuracy of general economic forecasts and on several related topics, including factors that affect forecasting procedures and performance; properties and quality of forecasting tools and targets (indicator series, anticipatory data, national aggregates of income and output); and the dependence of forecasts and their errors on types of economic change, in particular cyclical developments and turning points. This work is, as usual, in many ways a continuation of previous efforts: from the early (1929) report of Cox [6], through the NBER conference volumes of 1951 and 1955 [27], [5], to the subsequent studies by Christ [2], Okun [20], [21], Stekler [23], Suits [24], Theil [25], [26], and others, much has been done to develop and apply methods of evaluating various types of economic forecasts. But the recent National Bureau studies have put research in this area on a more comprehensive and systematic basis by collecting and analyzing a substantial body of data on economic forecasts of various types. The availability of these relatively rich and authenticated quantitative materials reflects the great proliferation of forecasting activity in the last two decades.

The materials analyzed in the early stages of the project consist in large part of predictions by business economists, who are in fact the source of the great majority of economic forecasts in the United States. The demand for forecasts of economic conditions increased greatly in the last two decades, judging from the expansion of the corresponding output, and business management clearly had a very large share in that demand. The preference of business has been for unconditional and, increasingly, for specific and numerical predictions. Forecasts by company economists are for the most part unaccompanied by explicit specifications of the methods or models used. In aiming at the comprehensive economic aggregates, such as gross national product (GNP) and its major components, they are presumably motivated by the working hypothesis that forecasts of these macrovariables are needed for predicting the microvariables of direct interest to a company (notably its own sales). The microforecasts, however, are generally not available and were not studied.

[1] These publications include [18], [29], [11], [15], and [4] (see References at end of this paper). These are reports of a study of short-term economic forecasting that was supported by grants to the National Bureau from the Whirlpool Corporation, General Electric Company, Ford Motor Company Fund, Relm Foundation, and the U.S. Steel Corporation, as well as by other funds of the National Bureau. A grant of electronic computer time to the National Bureau by the International Business Machines Corporation was used for some of the statistical analysis.

A continuing quarterly survey of general economic forecasts, based on a new questionnaire designed by the NBER, was initiated in November 1968 by the American Statistical Association; since then, its results are being regularly processed and analyzed by the National Bureau.[2] In the past year, steps were also taken to start a systematic collection of quarterly forecasts and related statistics for several major econometric models of the United States economy. In this second, current phase of the NBER research on forecasting, econometric models are a major subject for study, with primary attention being given to their short-term forecasting qualities and related properties, such as the ability to simulate cyclical developments in the economy and the effects of policy changes. Two products of this work are forthcoming. One, by Evans, Haitovsky, and Treyz [9], deals with the predictive performance of the quarterly Wharton and OBE models; the other, by Zarnowitz, Boschan, and Moore [34], with the cyclical and other properties of these models and the FRB-MIT-PENN model as revealed by various nonstochastic and stochastic simulations.[3] Further studies in this area are in progress.

In the next section of this report, I attempt to identify and review the main lines of investigation pursued in the National Bureau studies of economic forecasting and related topics. In the third section, some results of these studies are brought up to date and evaluated. In the last section plans for future research are discussed.

2. Directions of Recent and Current Research

2.1 Measures of the Accuracy and Structure of Forecasts

Whatever services the forecasts are expected to render to the user, they vary greatly, and are not easily defined by an outside observer; but the usefulness of most forecasts is surely in the first place a positive function of their accuracy. We therefore began by working on methods of assessing the

[2]Mrs. Charlotte Boschan and I share the responsibility for this work. The results are reported quarterly in press releases published in each successive issue of *The American Statistician* beginning in April 1969. The survey is carried out among members of the Business and Economic Statistics (B & E) Section of the ASA, with nationwide coverage; the participants are business, government, and academic economists whose professional work involves regular forecasting of the course of the economy. Ten major economic indicators are predicted for each of the four quarterly periods ahead, the base levels of the forecasts are specified, and questions are answered regarding the major assumptions and methods used and the probabilities attached to the expected changes in some key variables. For a description of the survey, see [32]; for an appraisal of the predictions for the first three quarters covered, see [33].

[3]For descriptions of the Wharton-EFU (Economic Forecasting Unit), OBE (Office of Business Economics, U.S. Department of Commerce), and FRB-MIT-PENN models, see [10], [14], and [22], respectively.

degree of predictive accuracy and applying them to the collected "judgmental" forecasts ([29] and [16]).

Accuracy is evaluated first with the aid of a battery of statistical measures of the closeness with which predictions approximate realizations. Arithmetic and absolute averages and standard deviations of forecast errors are employed as simple descriptive devices. Regressions of actual on predicted values and the decomposition of mean square errors provide estimates of bias and inefficiency. The measurement of accuracy in this absolute sense is followed by comparisons of actual forecast errors with errors resulting from alternative extrapolations of the time series concerned. There is a progression of these benchmark extrapolations from the simplest "naive models" to technically rather sophisticated autoregressive forms; but even the best of them make little or no use of economic theory and pose relatively few requirements in terms of data and computational operations. Thus, these comparisons with the benchmark models show what, if any, is the net contribution of the forecasts to the information about the future that can be obtained from other quicker and cheaper methods.

Besides being useful as yardsticks of predictive performance, extrapolations can help explain the implicit structure of judgmental forecasts on the plausible assumption that most of these (and other) forecasts rely, to some extent, on various types of extrapolation. In [16, sec. III], methods are developed to decompose forecasts and forecast errors into extrapolative and other (autonomous) components. Application to a few sets of business forecasts shows how this approach can be used to evaluate the relative importance of extrapolations in generating such forecasts, and the effects of extrapolation error on forecasting error. In [17], further efforts are made to infer from the available forecast data, and particularly their estimated extrapolative components, how forecasters may be revising their expectations in the light of their past errors. Of the linear models of adaptive behavior examined, the one that receives most support and attention involves forecast revisions which are a fraction of the current observed forecast error, the fraction being smaller for longer-term than for near-term predictions. This analysis is extended in [8] and applied to the problem of how forecasts of future spot rates of interest are formed in the market (the working hypothesis here being that these market forecasts are given by the forward rates implicit in the term structure).

2.2 The Effects of Data Errors and of the Time Span of Prediction

In [29] several factors are shown to affect strongly the absolute and relative accuracy of the forecasts assembled there. One is the lack of accurate information about the conditions prevailing at the time the forecast is made. The initial level from which the predicted change is measured must itself be predicted; and although these base values are estimated at a close range, they often contain significant erors which contribute to the errors of forecast.

This is closely related to the broader subject of the effects of errors in current and past data on forecasting accuracy. In [4], successive revisions of provisional GNP estimates are found to reduce the GNP data errors on most occasions, mainly by reducing errors arising from extrapolations of past benchmark values. The provisional estimates are themselves partly near-term predictions, and they have some characteristics frequently observed in "true" forecasts such as the tendency to underestimate increases and levels of GNP. The estimates for a year just ended are on the average substantially more accurate than the forecasts for a year ahead, but they are not much better than the forecasters' own estimates of the current or base levels (which are typically made about three to four months earlier). The use of preliminary rather than revised GNP data appears to account for more than one-third of the observed average errors of the annual GNP forecasts [3].

Accuracy tends to diminish steadily as the forecast span increases [29]. Business economists' forecasts of GNP and industrial production, for example, are typically better than various types of extrapolation over periods from one to three quarters ahead. Forecasts for four quarters or more ahead, however, are generally not superior to simple extrapolations of the recent trend. Year-to-year forecasts are on the whole more accurate than even the more refined autoregressive projections, but they can be viewed as having mean spans of little more than six months.[4] Decomposition of relative mean square errors, in [16], shows that the contribution to predictive efficiency of the nonextrapolative (autonomous) component of the forecast typically declines as the span lengthens. At the same time, trend projections become potentially more useful, but forecasts fail to take sufficient advantage of them and consequently deteriorate faster than the best of the extrapolations as the predictive span is extended beyond a few quarters. However, marginal errors of multiperiod forecasts show no systematic rise, so average errors increase less than in proportion to the lengthening of the span. They also increase much less than the errors in simple naive models (see [29]).

Consideration of ingredients of general economic forecasts helps to explain these findings. In addition to extrapolations of some kind, forecasters use relations between the series to be predicted and known lagged values or estimated current values of other variables; various indicators of major changes in aggregate economic activity and in anticipatory data such as surveys of consumer and investment intentions and government budget estimates; and, finally, their own judgments. Each of these potential sources of forecast is likely to deteriorate as the span of prediction increases. The

[4] The annual forecasts are generally made in the late autumn for the calendar year that is about to begin; if they score well in the first two quarters, their record will be moderately good for the year as a whole.

forecasting relations between time series involve various lags, but typically the relations weaken as the lags are increased. Most of the leading business cycle indicators have relatively short effective forecasting leads beyond which their usefulness declines (see [18] and [19]). The same appears to apply to anticipatory data and probably to informed judgments and estimates generally.[5]

2.3 Errors in Predicting Economic Growth

Other studies have suggested that forecasters tend to underestimate changes in the predicted series.[6] This would not be a type of systematic error that forecasters could or should avoid if it merely resulted from failure to predict random variations in the actual values. Indeed, forecasts that captured all but the random component of the change can be viewed as optimal, and they would necessarily be underestimates in the sense of having a smaller variance than that of the actual values. (This last property applies more generally to all unbiased and efficient predictions of change; see [16, p. 18]). However, underestimation becomes undesirable if it applies to longer cyclical movements, not just to short irregular variations; predictions that systematically understate high values and overstate low values of the series are inefficient (being correlated with their own errors) and can be corrected (potentially) (see [16] and [29]). Finally, forecasts with a tendency either to under- or overestimate the actual values of the given series (its "levels") contain a bias, which is usually regarded as a very objectionable error that should be eliminated as far as possible. But all this refers to population or "long-run" characteristics. In the limited samples of comparable predictions and realizations that are typically available, bias and inefficiency are difficult to ascertain, let alone measure and project with sufficient confidence; hence attempts to remove or at least drastically reduce such errors are often frustrated.

[5] Most data for business and consumer anticipations represent single-span forecasts or expectations, but the OBE-SEC quarterly surveys of investment intentions produce two series of anticipated business expenditures for new plant and equipment: the "first anticipations" (A_1) reported early each quarter for the next quarter and the "second anticipations" (A_2) reported at the same time for the current quarter. A_2 is on the average substantially more accurate than A_1 (see Okun in [28, p. 436], with references). Also, in regressions of purchase rates for new automobiles on consumer buying intentions and attitudes (Bureau of the Census and University of Michigan Survey Research Center data) better results are obtained in most cases with a six-month lag of purchases than with nine- or twelve-month lags (see Table 5-2 and text in [13]).

[6] See Franco Modigliani and Owen H. Sauerlander in [5, pp. 288–289] and H. Theil in [25, Chap. III–VI], among others.

Most forecasts examined do underestimate on the average the growth of the economy as measured by GNP. About two-thirds of the annual increases in GNP during the period 1952–63 were underpredicted; and the same applies to the concurrent changes in personal consumption expenditures, a still more smoothly growing aggregate which experienced no year-to-year decreases. The declines in GNP were less frequently underestimated. Changes in series that fluctuated more and grew less vigorously (e.g., gross private domestic investment in 1952–63) have been overestimated about as often as they have been underestimated. The same applies to decreases in all the major GNP expenditure components taken together; but these were just as frequently missed (though the proportion of these turning-point errors varies greatly among forecasts from different sources). As for increases, in this set of predictions they were underestimated nearly half the time and overestimated more than one-third of the time, while turning-point errors accounted for the remaining one-sixth of the observations [29, pp. 45–51].

These findings suggest that the observed "underestimation of changes" reflects principally a conservative prediction of growth rates in series dominated by upward trends. As this implies, the levels of such series also are generally underpredicted [16, p. 19].

2.4 Predicting Cyclical Movements and Turning Points

Predictive errors appear to be affected by the cyclical characteristics of the forecast period [29, pp. 27–30]. Thus, the underestimation of the increases in GNP is typically largest for the beginning of a recovery from a business recession, when the growth rates are particularly high. Later in the expansion, the increases are usually smaller, at least in relative terms, and the amount of underestimation is on the whole much less (and the same applies, consequently, to the associated target levels of the forecasts). Indeed, overestimation prevailed in one period of retardation (1962) and was frequent in another (1967). In contractions, the predicted levels are often too high, sometimes because the decline turned out to be larger than expected but mainly because the downturn was missed.

Annual forecasts of GNP had some success in predicting both the frequency and the timing of the turning points [29, pp. 51–59; 19, pp. 3–5].[7] They are certainly superior in this respect to extrapolations, which are by and large incapable of *signalizing* the business cycle turns and are instead apt either to "smooth" them out of existence or to reproduce them with lags. It is true that forecasts can and occasionally do predict turns which

[7]Much of the detailed information underlying the discussion in this and the three following paragraphs has not yet been published; it is contained in Victor Zarnowitz, "The Record of Turning Point Forecasts of GNP and Other Major Aggregates," NBER (draft manuscript).

then fail to occur, while trend projections avoid giving such "false warnings"; but the latter errors are infrequent and this disadvantage of forecasts is outweighed by their advantage of missing fewer actual turning points. A naive assumption that next year will always produce a turning point would avoid all errors of missed turns, but it would of course be a very poor one for predicting comprehensive aggregates such as GNP in an economy that tends to grow most of the time; the forecasts are far better than this benchmark model in that they make fewer errors of the false-signal type. They are also much better than the almost equally naive assumption that GNP is a series of random numbers and also better than the somewhat less naive model that would treat the annual *change* in GNP as random [19, p. 4].

However, these comparisons do *not* imply that the forecasters were able to predict the turns in the months ahead; they indicate only an ability to recognize the turns with relative promptness. Consider the forecasts for 1954 made in November or December of 1953: They showed 1954 as lower than 1953 in terms of GNP, i.e., 1953 to be a peak year. But by this time the midyear at which the turning point in annual data is conventionally dated was long over. Actually, the peak in the business cycle occurred in or near July 1953, and late in the year (at the height of the "forecasting season") the decline was widely recognized, though not necessarily as a cyclical contraction or recession.[8] In 1957 the peak again occurred shortly after midyear and forecasts made in late autumn or early winter had only to recognize a contraction then in process; in 1960, the peak probably came still earlier (the NBER reference month is May), but the recognition process was generally slower, reflecting the disturbing effects of the 1959 steel strike and the shallowness of the contraction. In each of these three episodes, the task of predicting annual changes was also made easier by the widespread expectation (which proved largely correct) that the contraction would be short and mild. Furthermore, the timing of the troughs was early enough for the forecasts made late in 1954, 1958, and 1961 to benefit from many indications that each of the respective contractions had already ended or was about to.

Forecasts made near the middle of the year for the next calendar year account only for one-tenth of all annual GNP predictions collected in [29]. Their record is a great deal worse than that of the end-of-year forecasts and not much better than guesswork as far as turning-point errors are concerned [19].

Forecasts made two or four times in a year for sequences of four to six quarters ahead are more relevant for an appraisal of turning-point errors. The

[8] See Rendigs Fels, "The Recognition Patterns of Business Analysts," in [11, Part I, p. 28].

reason is that there are more turns in quarterly than in annual series, and they can be dated with greater precision. There are relatively few directional errors associated with increases, but many associated with decreases (that is, missed peaks). Of 194 predicted quarter-to-quarter rises in GNP, 155, or nearly 80 per cent, coincided with actual rises; of 19 predicted declines, only 4, or 21 per cent, did so.[9] As this shows, forecasters know very well that increases prevail heavily among the quarterly changes of GNP, and they make increases similarly dominant in their predictions (in fact, somewhat more so). What the forecasters evidently do *not* know is how to anticipate correctly when the declines are coming. Predictions for sequences of two semiannual periods ahead (from different sources) have a slightly better record, but they too are consistent with the general conclusion that there is little evidence of forecasters' ability to detect the major cyclical reversals in the economy well in advance of the event.

Statements about the business outlook in the leading business and financial publications show a pattern in the neighborhood of business cycle peaks and troughs. The analysts "become increasingly aware of first the possibility, then the probability, and finally the certainty of a turning point" [11, Part I, p. 47]. In the three months preceding the event, expectations of a turn become slowly but clearly stronger and more definite, and the process of increasing recognition continues for several months after the actual turn (as dated *ex post* in the NBER reference cycle chronology). The recognition of the four business cycle troughs in the period 1948–61 has been on the whole faster than the recognition of the four peaks. Evidently, the analysts, like the majority of forecasters, were basically optimistic about the growth prospects of the economy and the short-lived nature of the contractions. Given this prevalent attitude and the historical course of events, the downturns proved to be more difficult to predict and recognize than the upturns.

The system of scoring reports on the economic outlook developed in [11, Part I] was applied in a companion study to the appropriate statements in the minutes of the regular meetings of the Federal Reserve Board's Open Market Committee (FOMC) in 1948–60.[10] The results are of major interest in view of the importance of the FOMC in determining the nation's monetary policy. They indicate that the Committee was a little more successful than the

[9] An alternative way of describing this property of the quarterly forecasts is this: of 43 actual declines in GNP, 39 or nearly 91 per cent were missed. It may also be noted that most of the predicted declines relate to the first two quarters ahead; beyond that, for the third to sixth quarters ahead, very few declines were anticipated and none at the right times.

[10] C. Elton Hinshaw, "The Recognition Pattern of the Federal Open Market Committee," in [11, Part II].

average of the published business outlook reports in recognizing and confirming (not in predicting ahead of the event) the cyclical peaks and troughs of the economy. The FOMC definitely avoided false signals more successfully than did the business analysts.

Before the date of a business cycle peak or trough, the estimated probability of the occurrence of a turning point was typically below 50 (on a 0 to 100 scoring scale) for both the business analysts and the FOMC [11, Parts I and II]. Only one or two months *after* the turn would the odds begin to favor slightly the affirmation of a cyclical reversal over its negation. The odds would then reach 3 to 1 (probability of 75) about three or four months after the turn and odds of 9 to 1 (probability of 90) about five or six months after. This record is in most cases somewhat worse than that of the quantitative business forecasts [11, Part I, pp. 45–46]. It confirms the general lack of demonstrated ability to predict major turning points in aggregate economic activity ahead of the event.

2.5 Forecasts of the Major Expenditure Components of GNP

Forecasts of total GNP are often substantially better, in the sense of having smaller percentage change errors, than the forecasts of most major GNP expenditure components from the same source [29]. The over-all predictions apparently benefit from partial cancellation of errors in predictions of the components. While this is definitely preferable to the opposite case of positively correlated and mutually reinforcing errors, any gross inaccuracies in the components must of course be seen as detracting from the quality of the forecaster's product, even if these errors happen to be largely compensating. However, there are some systematic factors that are likely to confer a relative advantage on the GNP forecasts. Certain methods of forecasting, such as the use of business cycle indicators or monetary variables, are concerned directly with measures of aggregate economic activity rather than with any component expenditures or sectors of the economy, and hence may yield better forecasts for total GNP. Moreover, there are strong equilibrating or shock-absorbing forces at work in the economy, which effectively localize many disturbances that impinge upon the different processes or sectors. As a consequence, GNP is a rather smooth series dominated by a pronounced growth tendency, and it is therefore easier to predict than most of its components, which are much more volatile.[11]

[11] This rule has important exceptions, however: the massive aggregate of personal consumption expenditures, which accounts presently for over 60 per cent of GNP, follows a smoother course than total GNP, owing to the stability of personal outlays on nondurable goods and services and despite the volatility of spending on automobiles and other durables. The total of state and local governments' purchases of goods and services also shows fewer and smaller relative deviations from trend than does GNP.

There is strong evidence that the more volatile a series, the more difficult is the prediction of its relative changes and the greater the probable forecast error [29, Chap. 4]. Thus, only total and nondurables consumption and state and local expenditures are predicted about as well as, or better than, total GNP, according to the business forecasts for 1953–63 (see footnote 11). The errors in predicting percentage changes in personal consumption are far smaller than those in forecasts of gross private domestic investment, whereas the errors for total government spending are intermediate: larger than those in consumption but much smaller than those in investment. Within aggregate consumption, the errors for durable goods exceed greatly those for either the nondurables or services. Within aggregate investment, the record for producers' durable equipment turns out to be worse than that for total new construction.[12]

Although the errors of consumption forecasts are smaller than those for the other major GNP components when measured in deviations of percentage changes, they are large relative to the errors of appropriate extrapolations. The consumption aggregates (except for durable goods) are smoothly growing series that could have been predicted very well by simple trend projections; and, indeed, the average errors of the latter have often been smaller than those of recorded consumption forecasts. It would seem, then, that there is much scope for potential improvement in these forecasts, through better utilization of the historical content of the series.[13]

Improvement appears to be much more difficult to come by, but is probably even more necessary, in forecasts of components of investment, particularly residential structures, changes in inventories, and net exports. Even though these series have relatively weak trends and strong cyclical and irregular movements, and hence cannot be very effectively extrapolated by any simple means, their forecasts have on the whole proved to be either just slightly better or worse than mere extrapolations of last levels or average

[12] Expenditures on producers' durables have been less stable than those on structures in the period here covered. Expenditures on housing, however, have fluctuated widely and behaved quite differently from other major expenditure categories, reflecting the countercyclical effects of financial factors (mortgage credit was scarce in several periods of advanced expansion and relatively abundant in some periods of late contraction and recovery).

[13] The more volatile the series to be predicted, the less can the forecast in general gain from extrapolation. It is quite reasonable that forecasters should make greater use of extrapolation where this promises to be more efficient. Apparently, such distinctions are in fact made, for we observe that, e.g., forecasts of plant and equipment outlays have relatively smaller extrapolative components than forecasts of consumption [16, pp. 30–31]. However, the latter forecasts are still deficient in their use of the extrapolative potential of the consumption series.

changes. The predictions of net inventory change are particularly important for an appraisal of the business outlook in times of mild economic fluctuations, during which the "inventory cycle" is widely believed to play a major role; but these predictions are also particularly unsatisfactory.[14]

Aggregation of short-term expectations of business concerns about their expenditures on plant and equipment results in better predictions of the total of such capital outlays than are available from the independently made global forecasts of business fixed investment. This can be inferred from the markedly lower accuracy of the annual investment forecasts collected in [29] that were made before the McGraw-Hill Survey of Investment Intentions compared to those made after it. Also, the OBE-SEC investment anticipations series show not only very high simple correlations with actual plant and equipment expenditures but also high partial correlations, holding the investment forecasts constant. In contrast, the investment forecasts add very little or nothing to the statistical explanation of the variance of actual expenditures after allowing for the high correlation of expenditures with anticipations: the partial correlations here are small and as often negative as positive. This applies to quarterly as well as annual predictions [30, pp. 30–33].

2.6 Forecasts of Production and the Price Level

Among variables other than GNP and its major components, the FRB index of industrial production represents a favorite target of business economists' forecasts. It is generally predicted with a degree of accuracy similar to that of the GNP forecasts from the same sources, at least for the annual and shorter forecasts: comparisons in terms of index numbers on a common base somewhat favor GNP; comparisons with benchmark extrapolations in most cases favor industrial production [29]. For longer spans, predictions of the FRB index often come out better than those of GNP, according to the yardstick of simple trend projections. Monthly data provide more current information than quarterly data, which gives an advantage to the industrial production forecasts. The latter suffer less from the underestimation-of-growth bias than the GNP forecasts do, but they are more adversely affected by turning-point errors. In terms of correlation of actual with predicted changes, the forecasts of both GNP and industrial production are definitely superior to all examined types of extrapolation,

[14] For example, in business predictions for 1953–63, the absolute average error of forecasts (in current dollars) was greater for net inventory change than for total government expenditures and the major components of fixed investment, even though the average magnitude of the inventory change is very small compared with the typical levels of these other variables [29, pp. 39–40].

including the relatively accurate trend projections and autoregressive models. This advantage of greater efficiency appears to be widespread, and it is by and large not offset by the disadvantage of greater bias that the forecasts often have relative to the best of the benchmark extrapolations [29, Chap. 6]. The advantage can be traced to the contribution of the autonomous components of the forecasts [16, sec. II].

Forecasts of general price movements, although still much less numerous than those of GNP and related variables, attract increasing attention in these times of strong inflationary tendencies. Forecasts of the Consumer Price Index in recent years (mainly 1957–67, though some sets go back to 1953 and before) have on the average been more accurate than the naive model predictions. Some but not all of them were also better than simple trend (average historical change) extrapolations. Forecasts of the Wholesale Price Index were on the whole considerably weaker, but this is partly due to the extraordinary stability of this index in the years 1958–64 (which favored the naive "same-level" model). Both the CPI and the WPI forecasts show generally small average absolute errors, relative to the levels of the indexes and to the errors of forecasts of other comprehensive series such as GNP; but the changes in the indexes were also relatively small and less than their counterparts for these other variables. Forecasters have tended to underestimate the large changes and to overestimate the small changes in the price indexes, which suggests a certain inertia of expectations; there is more uniformity and predictability in the predicted than in the actual price-level behavior. However, here too, there are definite positive correlations between the forecast and the realized changes, and mechanical extrapolations of past price levels or changes could not have done as well in this respect [31].

The effects of the price errors on the GNP forecasts are not always adverse. For several forecast sets, these errors were negatively correlated with the errors of the implicit quantity component of GNP.[15] As a consequence of partial offsets between the errors of the quantity and price components, forecasts of GNP in current dollars have been on the average more accurate than the implicit forecasts of GNP in constant dollars for most of the private sources of the CPI and WPI predictions assembled in [31].

[15]Since there are no predictions of the GNP price deflator from these sources, a weighted combination of CPI and WPI forecasts had to be used as a rough approximation. The resulting composite price-level forecasts were applied to the GNP forecasts from the same source to obtain the implicit predictions of GNP in constant dollars. Expressing all figures as forecasts of percentage changes, additive quantity and price components were then computed. As an outcome of this decomposition, three series of predictions are available for each of the forecast sets covered, relating to GNP in constant dollars, the composite price level, and GNP in current dollars.

Materials on business forecasts of different variables suggest that turning points in lagging indicators such as plant and equipment expenditures and consumer prices are on the whole more accurately predicted than turning points in leading or coincident indicators such as inventory change and wholesale prices. Forecasts of the laggers benefit from observation of related series that move earlier in the course of the business cycle [19, p. 5].

2.7 Econometric Model Forecasts and Simulations

Moore [19, pp. 12–14] presents an analysis of predictions of annual percentage changes in GNP in constant dollars, based on an econometric model that grew out of the early work by Lawrence Klein and Arthur Goldberger and was developed by Daniel Suits [24]. The forecasts are prepared by the Research Seminar in Quantitative Economics, and they are usually presented in the November preceding the target year at the Conference on the Economic Outlook, both at the University of Michigan.[16] Sixteen forecasts, for 1953–68, are evaluated in [19]. Their mean absolute error is about 1.5 percentage points, less than half of the corresponding mean of actual changes (3.8 percentage points). The results for 1953–63 are almost identical, and they resemble closely those obtained for one of the judgmental forecast sets in [29], which also refers to changes in real GNP during the same period.[17] These comparisons suggest that, in predicting changes in GNP excluding the price movements, forecasts with this econometric model "appear to stand up comparatively well" [19, p. 14]. Similarly, Cunnyngham's analysis of GNP predictions in current and constant dollars led to the conclusion that these forecasts "have been about as accurate as the better business forecasts" [7, p. 60].

[16]The predicted percentage changes used in [19] were computed from published reports of the Michigan Conference and data supplied by the Research Seminar. The forecasts were not produced by a single model, but rather by a whole family of models plus judgments about not only the inputs of exogenous variables but also various adjustments of the constants in some of the equations (assumptions of nonzero residuals). The models themselves were modified (mostly elaborated) almost from year to year. Sometimes alternative forecasts for a given year were made, based upon different data or policy assumptions; see [19, Table 5, p. 13] for the numerical identification of such forecasts and the selections made.

[17]The other forecasts in that collection (also covering the years 1953–63) are not directly comparable because they refer to GNP in current dollars and are therefore affected by errors in predicting the price levels. The mean absolute errors of forecasts of annual rates of change in current-dollar GNP varied from 1.4 to 3.0 and averaged 2.0 percentage points for the eight sets concerned (only one of these sets, however, had a mean error of more than 2.3 per cent). The mean of the coefficients of correlation between the predicted and actual percentage changes in GNP is 0.78 for this sample; the correlation for the Suits model is 0.72.

However, it is very important to keep in mind that these are *ex ante* predictions that involve judgmental forecast of the exogenous inputs, recurrent revisions of the model, and frequent adjustments of the constant terms ("fine tuning")—all of these decisions being subject to modifications in the light of the preliminary and other information available to the econometric forecaster. Such results, therefore, tell us something about how well the latter forecasts rather than about how well the econometric model forecasts. To answer questions about the forecasting quality of both the models and the econometrician's judgments, and about how these elements interact, it is necessary to analyze not only the *ex ante* but also the *ex post* forecasts (which use the actual historical values for the exogeneous inputs) and to take into account the adjustments of the constant terms.

A systematic analysis of this kind was made recently for the quarterly Wharton-EFU and OBE models in [9]. Its principal findings are very revealing. The true *ex ante* forecasts are superior to the *ex post* forecasts with the same constant-term adjustments and lagged values of the endogenous variables. Thus, surprisingly, the forecasts made with the model-builders' estimates for the exogenous variables are here better (mostly by substantial margins) than the forecasts with the correct values of these variables. The true *ex ante* forecasts (with subjective adjustments of constant terms) are generally much better than forecasts that use the same estimated exogenous inputs but either without adjustments or with only mechanical adjustments of constant terms. The "*ex ante*" forecasts in the latter category tend to be more accurate than the *ex post* forecasts of the OBE model with the same mechanical adjustments or with none, but this result is reversed for the Wharton model.[18]

It is clear that the intercept adjustments must have often improved these forecasts considerably, which speaks well for the econometrician's judgment qua forecaster. But the poor showing of the *ex post* forecasts indicates that "econometric models cannot generate good forecasts if they are used only in a mechanical fashion" [9, p. 160]. Indeed, such forecasts are in a great many cases worse than the simplest naive models of the "same change" or even of the "no change" variety, when made over the shortest spans (of one or two quarters) and without any adjustments. The accuracy of the forecasts relative to the naive models improves with the lengthening of the span to three to six quarters, which reflects the fact that the forecast errors (1) have relatively small systematic components, and (2) are mainly due to "imperfect

[18]The period covered by the Wharton model forecasts is I-1965–IV-1968; that covered by the OBE model forecasts is II-1967–IV-1968. The forecasts are quarterly, and each consists of a sequence of predictions for either one to six quarters ahead (Wharton) or for one to four quarters ahead (OBE).

covariation" [25, pp. 35–37], one can also say that the forecasts track the longer movements better than the shortest changes over the next quarters. However, since the errors of the *ex post* predictions are significantly large for the longer spans, and the naive models represent quite weak standards for such forecasts, none of these results can be viewed as really satisfactory.

Three different possible reasons are suggested in [9] for the finding that *ex ante* forecasts are better than *ex post* ones: (a) the *original* (subjective) adjustments succeeded in offsetting "bad guesses" about the exogenous inputs; (b) the actual data are at fault; and (c) errors in forecasts of exogenous variables tend to cancel the misspecification errors in the models. The authors dismiss (a) on the ground that, in almost half the cases, the *ex ante* forecasts retain their superiority even when no adjustments or the same *mechanical* adjustments are used in these and the corresponding *ex post* forecasts. They treat (b) rather more seriously, though viewing only the figures for government expenditures as likely to be distorted in reporting.[19] Finally, (c) is discussed in the form of the offsetting effects of underestimation of government spending and overstatement of fiscal multipliers.

Of these factors, (c) is most disturbing and (b) is not damaging at all as far as the predictive quality of a model is concerned. My own conjecture is that the role of (b) is likely to prove rather limited and that of (c) principal, while (a) could well be considerably more important than implied by its treatment in [9]. But certainly this whole subject is exceedingly interesting, and it deserves much further investigation.

Sample-period simulations of the Wharton model for I-1953–IV-1964 and of the OBE model for III-1955–IV-1966 are also presented and discussed in [9]. These are calculated for six quarters ahead from each starting date, that is, from each consecutive quarter in the sample period.[20] They use actual data for the initial conditions and the exogenous variables but model-generated lagged values of endogenous variables after the starting date

[19]The reason is that these expenditures and, in particular, defense spending may be entered in the national income accounts as inventory investment while the goods are being produced and as government purchases only when they are ready for delivery and paid for [9, pp. 138–139].

[20]For the Wharton-EFU model, the sample period actually starts in 1948, but the earlier years were excluded from these simulations because of the difficulties with the solution program encountered for some of the six-quarter intervals during the Korean War years. Toward the end of the period covered, in 1963–64, the intervals were gradually reduced to 5, 4, . . . , 1 quarters so that they would not reach beyond the sample period. For the OBE model, however, this was not done: the simulations do not stop in IV-1966 but continue for six quarters regardless of the originating date. This is said to be unlikely to affect the main conclusions more than slightly [9, pp. 34-35, 72].

over each simulation period. Their average errors increase steadily as the prediction span lengthens from one to six quarters, but they tend strongly to decline with the span when taken relative to the average errors of the naive model of last-known level projections. For the one-quarter span, the Wharton simulations are not much better than the extrapolations with this naive model. The simulations of the OBE model tend to have smaller errors than the simulations of the Wharton-EFU model for the shortest spans, but the differences narrow and become minor for the longest spans.

The results for the sample-period simulations are generally very different from those for the *ex post* forecasts beyond the sample period. The latter have much larger errors (which for the shortest spans are often greater than the naive-model errors, as noted before). Also, the constant-term adjustments, which are essential for obtaining reasonable *ex post* predictions in the forecast period, have little effect on the accuracy of the sample-period simulations.

Different types of simulations are examined in [34]: nonstochastic model solutions for the entire sample period and for six-quarter intervals around business cycle turns, and stochastic simulations for long (hundred-quarter) spans starting at the end of the sample period and extending into the future. The main concern in compiling this study is with the dynamic properties of the models (Wharton, OBE, and FRB-MIT-PENN) and their relation to the observed cyclical characteristics of the U.S. economy. This lies largely outside the boundaries of the subject of forecasting and is treated in a separate paper.[21] However, it should be stated here that the analysis in [34], like that in [9], indicates that the models, even though concerned largely with short movements over a few quarters, do not track such movements very well (judging, in [34], from the record of six-quarter simulations in the vicinity of business cycle peaks and troughs).

2.8 Standards of Predictive Performance

Even the least "naive" of the extrapolative models utilize only the past history (and usually only recent history) of the forecast series; they leave out much information of potential predictive value and should not represent very demanding standards for the short-term aggregate forecasts. Sets of past predictions that have proved superior to the benchmark extrapolations themselves provide a higher and more realistic standard. However, the future period may be either more or less difficult to forecast than the past, and it should be desirable to improve upon, not just match, the historical record.

[21] See Victor Zarnowitz, "Econometric Model Simulations and the Cyclical Characteristics of the U.S. Economy," another paper prepared for this colloquium.

Moreover, comparisons with what happened may over- or undervalue conditional forecasts, depending on the degree to which the conditions have or have not been met. For these stated reasons, a different standard is proposed and experimentally applied in [19, sec. IV]. A composite index of indicators that have generally led at business cycle peaks and troughs appears to anticipate GNP by six months: percentage changes between *fiscal-year* averages of that index are well correlated with percentage changes between the *subsequent calendar-year* values of GNP. This relation would yield predictions that are mechanically quite accurate but also late and of short range; to get a benchmark model that would match actual forecasts made, say, in October-November, the percentage changes in the index are calculated from the preceding fiscal-year average to the third quarter of the year preceding the calendar year being forecast.

The method yields the following mean absolute errors of relative change in GNP (in percentage points): 1.8 for 1953–63, 0.7 for 1962–67. The corresponding error figures are higher for both an average of eight sets of business forecasts from [29] (2.0 for 1953–63) and the forecasts prepared by the Council of Economic Advisers and published each year, beginning with January 1962, in the *Economic Report of the President* (1.3 for 1962–67). According to analogous measures for GNP in constant dollars, forecasts with indicators were slightly less accurate than forecasts with Suits' econometric model in 1953–67 and slightly more accurate than Suits' and the *Economic Report* forecasts in 1962–67.[22]

Benchmark forecasts with leading indicators are similar to *ex post* forecasts with econometric models, and particularly to predictions of GNP made from reduced forms of larger systems, where the coefficients represent weights applied to the predetermined variables so as to measure the net direct or indirect effects of these variables on GNP. Still better, they are analogous to forecasts from single equations which include only known exogenous variables as predictors. They share the general advantages and disadvantages

[22]The mean absolute errors (in percentage points) are: Indicators: 1.6 for 1953–67, 0.9 for 1962–67; Suits' model: 1.5 for 1953–67, 1.1 for 1962–67; *Economic Report:* 1.1 for 1962–67. It should be noted that the forecasts of the indicator model for 1962–67 are based on regression equations fitted to data for the prior period, 1949–61. The forecasts for 1953–63, however, benefit from being based on regression equations fitted to concurrent data (for 1952–67).

of such predictive equations vis-à-vis the larger systems.[23] Their specific source of strength is the selection of variables with strong anticipatory elements; but, let us recall, the leading indicators were chosen, not with regard to their correlation with subsequent GNP movements, but mainly for the consistency of their timing in the business cycle. The theoretical rationale for including some of the indicators in this particular synthetic form of relationship seems weak, and problems of various kinds arise at this point, notably with respect to differences in the aggregation levels, the timing and area of impact, the distribution of the lags involved, etc.; but it is not at all clear what effects these factors have on the proposed method and what can be done to reduce any difficulties they may cause.[24]

The single-equation and reduced-form approaches to forecasting have, of course, a great many different applications. The relative effectiveness of the

[23]Large models can accomodate more predictors, which is an advantage if the latter are properly selected and related to each other; but the chances of misspecification and propagation of errors are by the same token large, too. Advances in technical ability to handle such models have accelerated enormously with the progress in computer hardware and software, outstripping the growth in both the required theoretical knowledge and the availability of proper data. Smaller models suffer correspondingly less from such discrepancies, being simpler, more easily surveyed, and less demanding of the knowledge of economic relationships and the data needed to estimate them; but they may not be comprehensive enough to do the job well. Single-equation models represent an extreme class in this ranking by size and simplicity and are strictly applicable only to one-way cause-and-effect relationships, not to mutual dependencies. For *ex ante* predictions, they must employ known lagged values or outside forecasts of the independent (exogenous) variables; otherwise they can provide only *ex post* predictions conditional on the knowledge of the exogenous inputs.

[24]It is possible to establish by experimentation what is the best simple lag in the relation between the changes in the composite of the indicators and the subsequent changes in GNP (for some work along these lines, see [19, pp. 16–17]), and this could be extended to distributed lags. But what is optimal in the sample period need not be so in the forecast period, and the stability of the lag may be especially problematic here because we are dealing with composites of series with different timing properties.

It may also be instructive to experiment with indexes that differ in the number and identity of the indicators included (again, some results on this are reported in [19]), and likewise with different dependent variables (for example, with employment or industrial production instead of GNP). In short, the procedure is quite flexible and it is easily kept up to date, too. But these advantages derive from its basically mechanical nature. It is because of the latter that the method is treated as another benchmark model, though one that may not be easy to beat. This implies the expectation that, over time, skilled judgmental use of the indicators should yield better forecasts. (Analogously, judgmental adjustments are viewed in [9] as necessary means for obtaining good-quality forecasts from econometric models.)

leading indicators model clearly suggests a simple but important generalization: that in this comprehensive class of applied economic forecasts the use of predictors with early timing ("leading" series, "anticipations" data) has a comparatively high probability of success. Such predictors should therefore be in strong demand, and one might expect to see much effort spent on increasing their limited supply by data compilation, experimentation, and testing. The leading indicators and the anticipations data, although for the most part a by-product of more basic research, have indeed long been a subject of much scientific interest in their own right, in recent years primarily at the National Bureau and at the IFO-Institute in Munich; and there is little doubt that practicing forecasters use such materials widely. But such data are still not fully and systematically incorporated in formal econometric models, even where forecasting of short-term movements of the economy is a major objective. One probable reason for this is that their use imposes limitations upon the model. When an aggregate anticipations variable is included, which is based on microdata obtained from surveys of the relevant decision-making units, it is difficult to extrapolate the model successfully beyond the time span covered by the survey because the anticipations themselves cannot be well predicted. A similar problem exists for the leading indicators.[25] Nevertheless, there is surely both need and scope for making more and better use of such data in forecasting with econometric models (and indeed in short-term economic forecasting generally).[26]

3. Forecasts of the U.S. Economy: A Review of the Updated Record

In this section that follows, several standards of predictive performance are applied to updated materials on forecasts of various types, from "purely" judgmental to those made by or with econometric models, including comparisons between forecasts from the same sources for different periods; between forecasts by different persons and methods made for the same

[25] At least some of these early-moving series make much better tools than targets of forecasts. Thus, as shown in detail in Victor Zarnowitz, *Orders, Production, and Investment—A Cyclical and Structural Analysis* (New York, NBER, forthcoming), new orders precede output and shipments in manufacturing industries in which production to order is important, and they predict these activities as well as possible. However, such orders contain a large element of "autonomous" expectations and are themselves poorly predicted. (This is shown, in particular, by microdata on business forecasts of new orders for nonelectrical machinery in Machinery and Allied Products Institute, *Capital Goods Review,* December 1965.) In fact, it is where new orders cannot be predicted sufficiently well at acceptable costs that production will be largely "to order"; where they can be so predicted and supplying them appears profitable, production will be "to stock."

[26] Further remarks on this point are made in section 4 of this paper.

periods; between *ex ante* and *ex post* forecasts for some econometric models; and between forecasts and benchmark predictions or extrapolations.

3.1 Private Judgmental Forecasts of GNP, 1953–69

The record of business and other private judgmental forecasts presented in [29] ends in 1963. Since then some of these forecasts were discontinued or excluded, while a few new sets were added to our collection. The amounts of continuous data from the same sources are now sufficient for compiling a record of forecasts through 1969 and for attempting to answer the question: How does the forecasters' performance in the most recent years (1964–69) compare with their performance in the earlier postwar period (1953–1963 or parts thereof)?

Table 1 shows some summary measures of accuracy for eight different sets of annual forecasts of GNP relating to the years 1953–69. The sources are: a group of economists from various industries, government, and academic institutions; a graduate faculty group; the economic staffs of three large business firms in insurance, banking, and manufacturing; a business publication; an individual university researcher; and a business consulting service. Five of the sets were included in [29], three (K, L, and M) were not.

For the two groups (A and M), which are relatively small, averages of separate forecasts made by the members are used, and it should be noted that such averages tend to be more accurate over time than most of the forecasts by the individual participants in the given group because of compensating errors among the member forecasters [29, pp. 123–126]. Furthermore, comparability of the different forecasts sets is impaired by differences in the periods covered and in the dates of issue (late forecasts can take advantage of recent information not available for earlier forecasts). However, it is not the question of who the best forecasters are that concerns us here but rather the over-all accuracy and some other interesting properties of the forecasters' products.

Since the initial level from which a change is predicted is itself as a rule unknown and must be estimated, the error in forecasting the change typically differs from the error in forecasting the future level, the difference being equal to the error in the estimated current position or "base" of the forecast. This base estimate should always be reported by a forecaster but only too often it is not; however, in recent forecasts this deficiency seems to be much less frequent, perhaps because the need to remove it, stressed in the literature reviewed in this paper, is being increasingly recognized.[27]

[27]Where the base forecasts are not reported, they must be imputed to make possible an analysis of the forecasts of change. Such imputations were made for this study by extrapolative methods selected so as to approximate as closely as possible any base estimates for other periods made by the given source (clues of this sort are often available; where they are not, methods approximating the results for other, similar forecasts were used). Compare [29, pp. 32-35].

TABLE 1

Annual Forecasts of GNP: Average Errors of Prediction of Base Values, Changes, and Levels and Comparisons with Three Extrapolative Models, Eight Sets of Private Judgmental Forecasts, 1953-69

Forecast Set[a]	Period Covered	Mean Error (*ME*)[b] Base[d] (1)	Change (2)	Level (3)	Mean Absolute Error (*MAE*)[b] Base[d] (4)	Change (5)	Level (6)	Root Mean Square Error (M_P)[b] (7)	Ratios of Root Mean Square Errors: Forecast to Extrapolation[c] R_1 (8)	R_2 (9)	R_2^* (10)
A	1954–69	–3.7	–6.4	–10.2	4.4	10.6	12.7	14.9	.391	.761	.602
B	1953–69	–2.8	–3.2	–6.0	3.4	9.2	11.1	12.5	.328	.641	.504
C	1958–69	–3.4	–3.8	–7.2	3.7	8.7	10.7	11.8	.271	.642	.432
F	1953–69	–3.2	–4.4	–7.6	3.5	6.9	9.7	11.1	.290	.570	.448
G	1953–69	–2.9	–0.5	–3.4	3.4	7.7	9.6	11.9	.312	.614	.481
K	1957–69	–5.2	–1.2	–6.4	5.2	7.0	8.8	10.9	.261	.612	.413
L	1956–69	–4.5	–3.3	–7.8	4.5	7.4	9.6	12.1	.281	.699	.478
M	1953–69	–2.5	–6.8	–9.3	3.1	10.4	11.8	14.0	.374	.723	.562

[a]For a brief description of forecasts, see text.

[b]A forecast error (*E*) is defined as the predicted value minus the corresponding actual value ($P - A$). The actual values are the first estimates of the U.S. Department of Commerce for the preceding year, which appear in January. All entries in these columns are in billions of dollars. The formulas for the averages are: $ME = (1/n)\,\Sigma\,(P_t - A_t)$; $MAE = (1/n)\,\Sigma\,|P_t - A_t|$; and $M_P = \sqrt{(1/n)\,\Sigma\,(P_t - A_t)^2}$, where the summation is over all *n* periods covered by the given set of forecasts. The M_P measures (column 7) refer to level errors.

[c]The ratios are: $R_1 = M_P/M_{N1}$; $R_2 = M_P/M_{N2}$; and $R_2{}^* = M_P/M_{N2^*}$. The denominators, M_N, are the root mean square errors of three types of extrapolation: *N*1 refers to the projection of the last known level, *N*2 to that of the last known change, and *N*2* to that of the average historical change. See text for more detail.

[d]Errors in the estimated current positon or base of the forecasts. The base values are the forecasters' own estimates, except for sets A (before 1957), K, and L.

For the GNP forecasts, base errors all have, on the average, negative signs, as do the errors of change; accordingly, the mean errors of level, which equal the algebraic sums of the corresponding mean errors of base and of change, are also all negative and, when signs are disregarded, the errors of level are largest of the three (Table 1, columns 1–3). The errors are computed by substracting the actual from the predicted values; thus, given the upward trend in GNP, the negative signs of the mean errors indicate the forecasters' tendency to underestimate the increases and hence also the target levels of this variable. The minus signs of the base errors also are related to this tendency, which apparently prevailed for the preliminary official estimates of GNP as well.[28]

In terms of absolute averages, the level errors are in each case larger than the change errors, (the weighted means for all sets are $10.5 billion and $8.5 billion, respectively) and the base errors are the smallest (columns 4–6). The latter result accords with expectation, since the present or the recent past should be better known than the future. However, the base errors are certainly significant, amounting on the average to more than one-third of the corresponding level errors.

Comparisons of the forecasts (P) with benchmark extrapolations (N) were made in terms of the level errors only, by calculating the ratios of the root mean square errors (M_P/M_N). These measures involve squaring the individual forecast errors, so that greater weight is attached to large than to small deviations of predictions from actual values.[29] Three extrapolative models are used: $N1$, which projects forward the current level of the given variable (here, the base-year value of GNP); $N2$, which projects the last change; and $N2^*$, which projects an average historical change.[30] The ratios are labeled R and bear the subscript of the model used in computing the denominator, e.g., $R_1 = M_P/M_{N1}$.

[28]See section 2.2 above on the effects of data errors on forecasting errors, with references to [3] and [4].

[29]See Table 1, footnote b, on the definitions of the root mean square error of forecast M_P, and the other averages used: mean error *(ME)* and mean absolute error *(MAE)*. Except in the trivial case where all errors are equal, $M_P > MAE > ME$ (compare the corresponding measures for the level errors in columns 3, 6, and 7 of Table 1).

[30]The assumption of $N1$ is that $A_{t+1} = A_t + u_{t+1}$, where A is an actual value and u is a random error. Hence the forecast here is $P_{N1} = A_t^*$ (the preliminary estimate of the current value of the series). $N2$ specifies that $A_{t+1} = A_t + \Delta A_t + u_{t+1}$, and the corresponding forecast is $P_{N2} = A_t^* + \Delta A_t^*$. The assumption of $N2^*$ is that $A_{t+1} = A_t + \overline{\Delta A} + u_{t+1}$ where $\overline{\Delta A}$ is the average value of past changes in the given series as available to the forecaster from the historical record (for GNP, the starting date of that record is 1947). Hence, the forecast $P_{N2^*} = A_t^* + \overline{\Delta A}$.

As shown in the three last columns of the table, the ratios M_P/M_N are definitely less than 1 in all cases; that is, the root mean square errors are smaller for the forecasts than for any of the three extrapolative or naive models ($M_P < M_N$). For each of the forecast sets, R_1 is the smallest and R_2 the largest of the ratios, with R_2* ranking in the middle; simple averages of the ratios for the eight sets are 0.32 for R_1, 0.49 for R_2* and 0.66 for R_2*. This means that, of these models, *N*1 gives the weakest and *N*2 the strongest results for GNP in the years from 1953 through 1969. But these are all exceedingly simple models, which ought to be viewed generally as providing only minimal standards of predictive accuracy. Although models such as *N*2 and *N*2* can be relatively effective for series with pronounced trends which are approximately linear over the periods covered, it will be desirable in further work to confront the GNP forecasts with more stringent yardsticks such as autoregressive predictions with certain optimal properties.

3.2 Forecasts Before and Since 1964

In Table 2, the GNP forecast records summarized in the preceding section are analyzed for two separate subperiods: one beginning in 1953 or later and ending in 1963, the other including the years 1964–69. The table shows that the root mean square errors of the level forecasts for the latter period exceeded their counterparts for the former period in each case and often by relatively large margins (columns 1 and 5). For all eight sets, these M_P figures average about $10 billion in the periods through 1963, $15 billion in the years 1964–69, and $12 billion in all years covered.

However, larger absolute errors do not necessarily denote worse forecasts. The size of errors depends on the size (as well as dispersion) of the values to be predicted. In forecasts of two series with similar behavior, the errors for the series that has lower levels are likely to be smaller on the average (e.g., GNP in 1958 dollars is predicted with smaller absolute errors than GNP in current dollars). GNP measured in current dollars rose from $365 billion to $589 billion between 1953 and 1963 and from $629 billion to $932 billion between 1964 and 1969; the observed increase in the average forecast errors appears rather moderate when compared to the contrast between the average levels of GNP in the two periods. Indeed, the ratios R_1 are on the whole much smaller for the GNP forecasts relating to 1964–69 than for those relating to the pre-1964 years (Table 2, columns 2 and 6). Thus, comparisons with the naive "same-level" projections show the more recent predictions in better light than the earlier ones.

The ratios R_2* also are decidedly lower for the 1964–69 forecasts than for the pre-1964 ones (columns 4 and 8). It can be said, then, that the GNP forecast errors increased on the average less between the two periods than did

TABLE 2

Annual Forecasts of GNP: Comparisons with Three Extrapolative Models, for Periods Before 1964 and 1964–69, Eight Sets of Private Judgmental Forecasts

Forecast Set	Forecast Periods Through 1963[a]				Forecast Period 1964–69			
	Root Mean Square Error (M_P)[b]	Ratios of Root Mean Square Errors: Forecast to Extrapolation[c]			Root Mean Square Error (M_P)[b]	Ratios of Root Mean Square Errors: Forecast to Extrapolation[c]		
		R_1	R_2	R_2^*		R_1	R_2	R_2^*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A	12.5	.506	.651	.781	18.2	.331	1.219	.497
B	10.7	.435	.553	.699	15.2	.276	1.024	.421
C	11.0	.424	.534	.797	12.5	.226	0.840	.343
F	8.8	.359	.457	.578	14.3	.258	0.959	.394
G	7.9	.322	.410	.518	17.0	.306	1.140	.467
K	8.5	.342	.453	.656	13.2	.238	0.882	.363
L	7.2	.283	.359	.591	16.6	.299	1.114	.456
M	12.8	.531	.839	.841	15.8	.294	1.061	.436

[a]For B, F, G, and M: 1953–63; for A: 1954–63; C: 1958–63; K: 1957–63; L: 1956–63.

[b]In billions of dollars. All means refer to level errors. See Table 1, footnote b.

[c]See Table 1, footnote c.

the mean annual change in GNP measured cumulatively over the postwar years.

In contrast, the R_2 ratios present a picture that is clearly unfavorable to the more recent forecasts in all but two sets (columns 3 and 7). For 1953–63, these ratios are concentrated in the 0.4 to 0.6 range; for 1964–69, in the 0.8 to 1.1 range. In periods of steady growth without turning points and major retardations, such as the middle and late sixties, the N_2 extrapolations of last change in GNP have a great advantage; in times of weaker trends and stronger cyclical movements, such as the post-Korean decade, this model performs much worse.

Inspection of the underlying series of individual forecasts and their errors shows that underestimation of growth was much more pronounced in 1964–69 than in the earlier period. The average year-to-year change in GNP (disregarding sign) was, $22.2 billion in 1953–63 and $55.6 billion in 1964–69; the corresponding averages for the predicted change are $21.7 billion and $46.5 billion.[31] The annual forecasts of GNP for 1953–63 include 43 underestimates, 25 overestimates, and 7 turning-point or directional errors (among the 75 forecasts included, six relate to declines in actual values, and they include three under- and three overestimates). The forecasts for 1964–69 include 41 underestimates and 7 overestimates; no actual declines occurred, and none were predicted, so the class of turning-point errors is here empty. These results are consistent with those obtained from a different sample of pre-1964 forecasts in [29]: as reported in section 2.3 above, underestimation has been related primarily to increases in series with dominant upward trends.

The rankings of the forecasters according to the average size of errors as measured by M_P and according to any of the R ratios are very different in the two subperiods. The rank correlations based on the entries in the corresponding columns (1 and 5, 2 and 6, etc.) of Table 2 are all virtually zero.[32] In terms of these ranks, sets A, G, and L have slipped, the last two considerably, and sets K and (especially) C and M have gained better relative positions (for B and F, no determinate shifts in ranks are observed). These

[31]These averages cover all eight forecast sets included in Table 2 and are weighted by the numbers of observations in each set. The averages for the actual change include the actual values that correspond to all these predictions and are analogously weighted.

[32]The ratios M_P/M_N have a common denominator for all the forecasts in a given period (it is a certain specific value of M_{N1} for the ratios R_1, of M_{N2} for the ratios R_2, etc.). Ranking the forecasts according to any of the ratios must, therefore, yield the same results as ranking according to M_P, if the periods covered are the same. In the present case, however, the periods differ, and hence the rank correlations also differ slightly. The Spearman rank correlation coefficients are as follows (the subscripts refer to the columns of Table 2): $\rho 15 = -0.047$; $\rho 26 = +0.024$; $\rho 37 = +0.024$; and $\rho 48 = -0.190$.

findings tend to agree with those reported previously on the apparent lack of consistency in forecasters' performance from year to year [29, pp. 130–132].

Table 3 shows some measures of absolute and relative accuracy for selected forecasts of the major expenditure components of GNP, using errors computed by taking differences between the predicted and the actual percentage changes. The dollar levels of these series differ greatly and errors of forecasts relating to these levels, or to the corresponding dollar changes, lack all comparability between such variables. An analysis in terms of relative change errors is more appropriate.

Although the level errors are on the average larger for the 1964–69 GNP forecasts than for the pre-1964 forecasts, as shown in Table 2, the opposite applies to the relative change errors (Table 3, forecast sets B, C, and F for GNP, columns 1 and 4). For the investment and government expenditure forecasts, too, the relative-change errors in the earlier years tend to exceed those in 1964–69, while for the consumption forecasts the errors in the two subperiods differ little (columns 1 and 4).

The average errors in predicting year-to-year percentage changes are much smaller in forecasts of personal consumption expenditures than in forecasts of gross private domestic investment (GPDI) in either subperiod and in the total period covered. The errors for the residential construction component are even larger than those for total GPDI, while the errors for government expenditures tend to exceed the errors for consumption but are also much smaller than those for investment (columns 1, 4, and 7). These results parallel the findings in [29, pp. 36-40].

With a few exceptions, the R_1 and R_2 ratios in Table 3 fall below unity, indicating that these forecasts are more accurate than the corresponding naive-model projections.[33] For consumption and government expenditures, the $N1$ model generally works worse than the $N2$ model (i.e., $R_1 < R_2$), while for GPDI the contrary is true. For residential construction, $R_1 > R_2$ in the earlier years and $R_1 < R_2$ in 1964-69, but the differences are relatively small. The ratios, R_1 are in virtually all cases larger for the investment variables (particularly residential construction) than for consumption and government expenditures; the R_2 ratios are in nearly two-thirds of the comparisons larger for the investment variables.

From the evidence of the R_1 ratios one would infer that these forecasts have on the whole improved in recent years, but the R_2 ratios present a mixed picture: for consumption and GPDI, they are higher in 1964-69 than in earlier years; for the housing and government expenditures, they are lower.

[33]Three ratios for residential construction forecasts and one for a government expenditures forecast are larger than 1. These cases include two R_1 and two R_2 ratios (see Table 3, lines 10 and 15).

TABLE 3

Annual Forecasts of Relative Changes in GNP and Four of its Major Expenditure Components: Comparisons with Two Extrapolative Models for Periods Before 1964 and 1964–69, Three Sets of Private Judgmental Forecasts

	Forecast Periods Through 1963[a]			Forecast Period, 1964–69			Total Forecast Period		
Forecast Set	$(M_P)^b$ (1)	R_1 (2)	R_2 (3)	$(M_P)^b$ (4)	R_1 (5)	R_2 (6)	$(M_P)^b$ (7)	R_1 (8)	R_2 (9)
				Gross National Product					
B	2.15	0.395	0.531	1.70	.225	0.804	2.00	0.320	.576
C	2.04	0.374	0.505	1.31	.173	0.618	1.71	0.442	.531
F	1.70	0.312	0.419	1.28	.169	0.604	1.54	0.246	.443
				Personal Consumption Expenditures					
B	1.52	0.318	0.528	1.49	.205	0.785	1.51	0.263	.583
C	1.44	0.310	0.528	1.44	.198	0.759	1.44	0.236	.612
F	1.69	0.354	0.588	1.14	.156	0.601	1.52	0.264	.592
				Gross Private Domestic Investment					
B	12.83	0.901	0.569	5.90	.620	0.607	10.93	0.789	.582
C	8.50	0.570	0.313	3.98	.419	0.410	6.64	0.532	.637
F	7.20	0.506	0.318	2.95	.311	0.304	5.95	0.430	.314

Table 3 (concluded)

Forecast Set	Forecast Periods Through 1963[a] $(M_P)^b$ (1)	R_1 (2)	R_2 (3)	Forecast Period, 1964–69 $(M_P)^b$ (4)	R_1 (5)	R_2 (6)	Total Forecast Period $(M_P)^b$ (7)	R_1 (8)	R_2 (9)
				Residential Construction					
B	14.12	1.229	1.049	6.14	.538	0.593	11.92	1.032	.958
C	9.04	0.865	0.770	6.55	.575	0.632	7.89	0.714	.711
F	9.15	0.931	0.832	6.45	.556	0.623	8.10	0.762	.757
				Total Government Expenditures					
B	2.82	0.406	0.467	1.40	.150	0.397	2.45	0.279	.464
C	2.17	0.319	0.690	1.30	.140	0.368	1.79	0.219	.535
F	2.15	0.309	0.355	4.56	.489	1.292	3.21	0.364	.609

[a]For B and F: 1953–63; for C: 1958–63.

[b]In percentage points. All measures refer to relative change errors. See Table 1, footnotes b and c, for the general definitions of the root mean square error M_P and the ratios $R_1 = M_P / M_{N1}$ and $R_2 = M_P/M_{N2}$.

In sum, the record does not show that the assembled forecasts generally have either improved or deteriorated in the recent years. Comparisons of level forecasts limited to some of the benchmark models ($N1, N2^*$) would have suggested improvement, but this merely reflects the fact that these models themselves are less effective for GNP in the mid- and late sixties than in the preceding decade; the opposite applies to $N2$ and when this model is used as a yardstick the forecasts appear to have worsed. The average errors in predicting annual percentage changes have become smaller for the forecasts of GNP and some of its major expenditure components, both absolutely and relative to the corresponding errors of the $N1$ model; but again the comparisons with $N2$ suggest in large part the opposite. Thus the results vary for the different forecast sets, variables, and criteria. One can merely speculate about why a more general and definite improvement has not been achieved.[34]

3.3 Some Comparisons with Governmental Forecasts of GNP, 1962–69.

Forecasts of annual percentage changes in GNP and the general price level prepared by the Council of Economic Advisers (CEA) can be taken or inferred from the *Economic Report of the President* for the years since 1962. Moore has reconstructed these important predictions in [19]. In Table 4, the CEA forecasts of GNP are compared with the actual values and with some selected private predictions for each of the successive years 1962–69.[35]

The CEA forecasts underestimate the actual percentage changes of GNP in each year except 1962 and 1967, and so do the average private forecasts (represented by mean predictions for the eight sets used in Tables 1 and 2). The underestimation errors are smaller in the CEA forecasts than in the

[34]The expectation that forecasts should improve is based in the first place on the general view of forecasting as a cumulative learning process. However, some practitioners take the skeptical position that changes in the economy are often such as to make the experience gained in the period past a weak guide to the future. Since the economy followed a more stable upward course in 1964–69 than in the preceding decade, which included three business recessions, one might contend that the recent years were easier to predict. But it can also be argued that excessive reliance on lessons of the previous period of economic hesitancy has contributed to the conservativeness of so many forecasts in the more recent years of stronger growth and inflation.

[35]The CEA forecasts listed in Table 4, line 2, are updated figures from [19]. The actual values in line 1 are computed from the first Commerce estimates that are also used elsewhere in my analysis of the accuracy of GNP forecasts. They differ but slightly from the actual values used by Moore in [19], which are the first official estimates given in the President's *Economic Report.* The divergencies never exceed ± 0.1 percentage points, and the mean of the actual percentage changes, disregarding sign, is the same according to either set of figures.

TABLE 4

Forecasts of Annual Percentage Change in Gross National Product, President's Economic Report and Some Selected Private Sources, 1962–69

	1961–62 (1)	1962–63 (2)	1963–64 (3)	1964–65 (4)	1965–66 (5)	1966–67 (6)	1967–68 (7)	1968–69 (8)	Mean 1962–69 (9)	Absolute Mean 1962–69 (10)
Actual and Predicted Percentage Change										
1. Actual[a]	6.8	5.5	6.6	7.6	8.6	5.5	9.0	7.6	7.2	7.2
2. CEA[b]	9.4	4.4	6.5	6.1	6.9	6.4	7.8	7.0	6.8	6.8
3. Average, 8 forecast sets[c]	8.0	4.1	5.8	5.7	7.0	6.4	7.6	6.8	6.4	6.4
Errors in Forecasts of Percentage Change[d]										
4. CEA[e]	+2.6	-1.1	-0.1	-1.5	-1.7	+0.9	-1.2	-0.6	-0.34	1.21
5. Average, 8 sets[f]	+1.2	-1.4	-0.8	-1.9	-1.6	+0.9	-1.4	-0.8	-0.72	1.25
6. Set K	-1.0	+0.6	-0.7	-1.6	-0.7	+2.3	-0.5	-0.8	-0.30	1.02
7. Set F	+1.9	-1.5	-0.8	-1.4	-2.2	+0.0	-0.8	-1.3	-0.76	1.24
8. Set C	+1.9	-2.0	-1.9	-2.1	-0.5	+0.5	-0.7	-1.1	-0.74	1.34

[a]Based on the first estimates of the U.S. Department of Commerce for the preceding year, which appear in January.

[b]Forecasts by the Council of Economic Advisers. Source: *Economic Report of the President*, January 1962 through February 1969 (based on figures on dollar levels, dollar changes, or percentage changes as given in the *Report*).

[c]Includes forecast sets A, B, C, F, G, K, L, and M, each of which provides a full coverage of the years 1962–69. Each entry in columns 1–8 represents the mean of the eight forecasts for the given year.

[d]Errors equal predicted minus actual percentage change, in percentage points.

[e]Errors of the CEA forecasts that are listed in line 2. See footnote b, above.

[f]Errors of the average forecasts that are listed in line 3. See footnote c.

average private forecasts in all but one year; on the other hand, the overestimate for 1962 is considerably larger in the CEA forecast (Table 4, lines 1–3). The average errors disregarding sign show only a minute difference in favor of the CEA predictions, which may or may not be significant (lines 4 and 5, with the summary figures in column 10).

Of the eight private forecast sets, K has a somewhat lower mean absolute error than the CEA forecasts, and two sets F and C show errors that are on the average not much larger (lines 6–8). Set K also underestimates the rates of growth in GNP a trifle less than the CEA forecasts, but the other two sets show stronger underestimation tendencies (as shown by the size of the negative mean errors in column 9). Of the five sets not individually included in Table 4, all have larger mean absolute errors (ranging from 1.4 to 1.8 percentage points), and in only one is the underestimation lower than that of the forecasts that are included.[36]

Comparisons with the three simple benchmark models, *N*1, *N*2, and *N*2*, indicate that the forecasts represented in Table 4 are all definitely superior to such extrapolations. The average errors disregarding sign are 7.1, 1.9, and 3.6 percentage points for the *N*1, *N*2, and *N*2* projections, respectively.[37]

Finally, the examination of all available annual forecasts of GNP suggests that the variability of predictions from year to year tends to be greater than the variability of predictions by different forecasters in a given period. I suspect that this is indeed often so and not confined to this particular variable and period, for the following reasons (which, although plausible, are also difficult to test): For all their diversity, which is considerable, forecasts for a given macroeconomic variable in a given period must have much in common because competent forecasters use to a large extent the same data and similar methods, are exposed to the same major current events and prevailing expectations, and possibly influence each other directly. Hence one might expect many forecasters to share at one time in a common success and at

[36]The predictions in set G consist of four overestimates and four underestimates (in 1965–69) and have a mean error of –0.21 and a mean absolute error of 1.38 percentage points.

[37]The mean errors, listed in the same order, are –7.1, –0.6, and –3.6 percentage points. The *N*1 criterion is, of course, an extremely weak one for a series growing as fast and as steadily as GNP in the 1960's, and it is merely noted here as one of the conventional yardsticks. The *N*2* benchmark model used to be rather effective for the earlier postwar period; when applied to forecasts concentrated in the years 1953–63 this model produced generally more accurate projections than either *N*1 or *N*2 [29, pp. 83–90]. The deterioration in the performance of *N*2* in the more recent years is clearly due to the increased growth rates in GNP as compared with the earlier postwar period. The relative stability of the growth rates in the period covered naturally works strongly in favor of the *N2* model.

another in a common failure (using these terms in a broad and relative sense). As for the variability over time of economic forecasts and their errors, this reflects mainly economic instability which has many forms and causes; some types of economic change are more difficult to predict than others and different types prevail at different times.

3.4 Some Multiperiod Forecasts by Business Economists

Some of the collected forecast sets provide chains of predictions made at a given date for two or more successive periods. Thus set A includes forecasts made annually for two and four quarters ahead; set C, forecasts made at somewhat irregular intervals four times per year for spans ranging from one to five quarters; and set G, forecasts made semiannually for spans varying from one to six quarters. Table 5 shows the root mean square errors of the GNP predictions from two of these sources and also the ratios of these Mp figures to the corresponding statistics for the naive models $N1$ and $N2$.

The Mp figures virtually always increase with the span of the predictions in any of the periods covered, i.e., since 1964 and before (there are no exceptions of this rule in the cases illustrated in Table 5, columns 1, 4, and 7). The *increments* of Mp in the successive quarters vary considerably, however, and show no tendency either to rise or fall; so the marginal errors are not necessarily larger for the more distant quarters than for those in the nearest future and may even be occasionally smaller. To put it differently, the increase in the average errors is typically less than proportional to the increase in the forecast span.

The R_1 ratios are all smaller than 1 and are concentrated in the range between 0.4 and 0.6 for any of the spans and periods covered (but they are, in most cases, somewhat larger in the period since 1964 than before; see columns 2, 5, and 8). This means that all these forecasts are better (approximately twice as accurate on the average, in terms of the root mean square errors of the level forecasts) than the same-level projections of the benchmark model $N1$.

The R_2 ratios also are smaller than 1 for all included forecasts that refer to the periods before 1964 and the total periods covered, but they are throughout appreciably larger than 1 for the forecasts relating to the 1964–69 period (columns 3, 6, and 9). This confirms that the last-change projections with the naive model $N2$ present a rather difficult standard of comparison for the GNP forecasts in the second half of the 1960's (see the high R_2 ratios for the annual GNP predictions of 1964–69 in Table 2, column 7).

As the span of forecast is extended, the ratios R_1 and R_2 decrease more often than increase, but the relationships are not very strong or regular. For

TABLE 5

Two Sets of Forecasts of GNP over Spans of from One to Five Quarters, Comparisons with two Extrapolative Models for Periods Before 1964 and 1964–69

Span of Forecast (no. of quarters)	Forecast Periods Through 1963[a]			Forecast Period 1964–69			Total Forecast Period[b]		
	$(M_P)^c$ (1)	R_1 (2)	R_2 (3)	$(M_P)^c$ (4)	R_1 (5)	R_2 (6)	$(M_P)^c$ (7)	R_1 (8)	R_2 (9)
				Set C (1958–69)[d]					
1	7.5	.412	.547	10.9	.732	1.361	9.0	.539	.767
2	10.8	.456	.507	17.9	.585	1.910	14.2	.531	.811
3	12.3	.413	.413	22.5	.512	1.668	17.2	.488	.702
4	15.0	.417	.414	27.1	.431	1.569	20.5	.425	.672
5	19.5	.428	.368	28.9	.358	1.312	24.3	.379	.588
				Set G (1953–69)[e]					
1	9.4	.507	.497	9.3	.502	1.371	9.4	.505	.606
2	10.2	.402	.366	14.0	.460	1.638	11.3	.431	.520
3	12.4	.418	.372	21.8	.466	1.440	16.4	.448	.566
4	13.8	.389	.324	25.6	.411	1.428	18.9	.409	.554

[a]For C: 1958–63; for G: 1953–63.

[b]For C: 1958–69; for G: 1953–69.

[c]In billions of current dollars. All measures refer to level errors. See Table 1, footnotes b and c, for the general definitions of the root mean square error M_P and the ratios $R_1 = M_P/M_{N1}$ and $R_2 = M_P/M_{N2}$.

[d]Includes 22, 20, 19, 13, and 7 forecasts for spans of one to five quarters, respectively, in the period 1958–63; 19, 14, 13, 8, and 6 forecasts for the same spans in the period 1964–69.

[e]Includes 20 forecasts per span for 1953–64. In the period 1964–69, includes 12 forecasts for spans of one or two quarters and 11 forecasts for spans of three or four quarters.

all observations in Table 5, declines in the ratios prevail over rises 30 to 12, and for the largest samples (total forecast periods) the dominance of the declines is even more marked (11 to 3). Hence it appears that the forecasts tend to *improve* relative to both naive models as the span lengthens.

The quarterly and multiperiod forecasts for the recent years, like the annual forecasts, make a better showing when evaluated in terms of errors of relative change instead of level errors (compare the measures for the annual GNP predictions in Tables 2 and 3). In the course of the present study, the absolute and relative accuracy of the updated multiperiod forecasts of relative changes in the major expenditure components of GNP was examined, but only a brief statement about the general results of this analysis is warranted here.[38]

The average errors of forecasts of consumption expenditures, gross private domestic investment, government spending, and several components of these categories, are all found to increase with the span of prediction, for the most part monotonically. However, the marginal errors of these forecasts do not increase systematically, that is, the average errors do not increase as fast as the span of forecast. The R_1 and R_2 ratios become in most cases lower as the forecasts reach further out into the future, indicating that the longer-range predictions are more accurate relative to the naive models than the short ones.[39] All of this agrees with, and amplifies, the findings obtained for the GNP forecasts, as summarized above.

Relatively few of the percentage change forecasts examined have root mean square errors exceeding those of the naive models (i.e., R_1 , $R_2 > 1$), even in the 1964–69 period. Exceptions are the forecasts of residential construction and net change in inventories, for which a large proportion of the R_1 ratios are larger than unity. Generally, $N1$ is more effective than $N2$ for these variables and gross private domestic investment as a whole (that is, R_1 tends to exceed R_2 here), while the opposite applies to consumption and government expenditures.

[38] The reason is that this material comes largely from one forecast set that is rather unlike the others in our collection of private judgmental forecasts, namely, set C, which evolved from a more informal forecast toward one that relies importantly on relationships estimated with an econometric model. These results, though interesting, are therefore somewhat limited in scope; yet it would take considerable space to present them in a more specific form.

[39] Changes in the R_1 ratios associated with successive one-quarter increases in the forecast spans consist of decreases in nearly two-thirds of the observations, while for the R_2 ratios the over-all proportion of decreases is over three-fourths. This includes the results for both the subperiods and the total forecast periods; when only the latter are considered, the connection of increasing spans with declining R ratios becomes more pronounced, particularly for R_2.

3.5 *Ex Ante* and *Ex Post* Forecasts with Two Econometric Models

Recently, Haitovsky and Treyz have brought up to date and extended the analysis of the forecasting properties of two econometric models, Wharton-EFU and OBE, which was first presented in [9]. I owe the data used in Table 6 to their courtesy.

The mean absolute errors of the *ex ante* forecasts *(MAE-XA)* for eight selected variables show the usual tendency to increase with the predictive span, from one to four quarters. There are only a few exceptions to this rule among the measures for the Wharton model (column 1), and none among those for the OBE model (column 4). However, the successive increments in these errors do not increase systematically, that is, the errors in predicting the first quarter are not always smaller on the average than the errors in predicting the second quarter, etc. In other words, the errors do not double and triple as the span of forecasts doubles and triples; in fact, they most often increase less than in proportion to the span, although no uniform and regular relationships emerge between these variables. All this agrees qualitatively with the findings for the judgmental forecasts as noted in [29, Chap. 5] and earlier in this paper.

The forecasts in question are the authentic *ex ante* predictions that involve judgmental adjustments of the constant terms by the Wharton and OBE forecasters. Several alternative mechanical adjustments were applied to the two models in [9] and again the updated version of [9]. The forecasts obtained with these reproducible adjustments tend to be less accurate than the true *ex ante* forecasts, often by substantial margins. Without any adjustments at all, the forecasts turn out to be still worse.

Ratios of the *MAE-XA* figures to the mean absolute errors of the simplest naive-model projections (*N*1 or *N*2, whichever proved more effective for the given variable) are listed in Table 6, columns 2 (Wharton-EFU) and 5 (OBE). For GNP in current and constant dollars, the implicit price deflator, and consumption expenditures on durable goods other than automobiles, *N*2 gave better predictions, hence the ratio shown is *MAE-XA*/*MAE-N*2. For the other four variables——expenditures on automobiles and on plant and equipment, net change in inventories, and the unemployment rate——*N*1 worked better and the entries correspondingly refer to the ratio *MAE-XA*/*MAE-N*1.

For both Wharton and OBE, the ratios for the forecasts of GNP in constant dollars are generally lower than those for current-dollar GNP (lines 1–10). The Wharton forecasts for GNP had larger errors than *N*2 (i.e., the ratios exceed 1) for the two shortest spans; the OBE forecasts had the larger errors for all but the shortest span. These poor results are apparently to a large extent due to very unsatisfactory price-level forecasts. Most of the ratios for GNP in 1958 dollars are smaller than 1 but several are not much smaller

TABLE 6

Wharton and OBE Model *Ex Ante* Forecasts of Eight Selected Variables over Varying Spans, Comparisons with Naive Model Extrapolations and *Ex Post* Model Forecasts, 1966–69

Span of Forecast (no. of quarters)	Wharton-EFU Model[a]			OBE Model[b]		
	MAE-XA[c] (1)	*MAE-XA*[d] / *MAE-N* (2)	*MAE-XA*[e] / *MAE-XP* (3)	*MAE-XA*[c] (4)	*MAE-XA*[d] / *MAE-N* (5)	*MAE-XA*[e] / *MAE-XP* (6)
			GNP (N2; billions of current dollars)			
1. One	3.8	1.05	0.63	2.1	0.68	0.49
2. Two	7.8	1.04	0.78	7.8	1.07	0.89
3. Three	10.4	0.82	0.89	12.7	0.99	0.93
4. Four	12.6	0.71	0.89	19.4	1.21	1.01
5. Next year	6.6	0.61	0.76	11.5	1.21	1.03
			GNP (N2; billions of 1958 dollars)			
6. One	2.8	0.91	0.56	1.8	0.64	0.77
7. Two	5.7	0.96	0.68	4.8	0.83	1.04
8. Three	7.2	0.71	0.68	7.0	0.64	0.95
9. Four	6.7	0.46	0.55	9.4	0.66	0.93
10. Next year	5.3	0.62	0.66	6.1	0.78	1.06
			GNP Implicit Price Deflator (N2; 1958 = 100)			
11. One	0.7	1.14	1.07	0.3	1.65	1.22
12. Two	1.0	1.17	1.27	0.6	1.78	1.33
13. Three	1.3	1.21	1.49	1.1	1.68	1.36
14. Four	1.8	1.40	1.62	1.5	1.89	1.30
15. Next year	1.1	1.22	1.38	0.9	1.91	1.41
			Consumption Expenditures–Automobiles (N1; billions of current dollars)			
16. One	1.5	1.13	0.89	1.6	1.23	0.91
17. Two	2.0	0.94	1.00	2.3	0.96	0.87
18. Three	2.8	0.96	1.00	3.2	0.88	0.84
19. Four	2.9	0.71	0.86	4.5	0.88	0.86
20. Next year	2.0	0.82	0.98	3.0	0.89	0.85

(continued)

Table 6 (concluded)

Span of Forecast (no. of quarters)	Wharton EFU Model[a]			OBE Model[b]		
	MAE-XA[c] (1)	MAE-XA[d] / MAE-N (2)	MAE-XA[e] / MAE-XP (3)	MAE-XA[c] (4)	MAE-XA[d] / MAE-N (5)	MAE-XA[e] / MAE-XP (6)
Consumption Expenditures–Durables, Nonautomotive (N2; billions of current dollars)						
21. One	1.4	1.36	1.07	0.7	0.83	1.00
22. Two	1.6	1.07	1.10	0.7	0.53	0.99
23. Three	2.2	1.47	1.15	1.1	0.86	1.09
24. Four	2.6	1.04	1.03	1.7	0.84	1.15
25. Next year	1.9	1.30	1.21	0.8	0.72	1.19
Investment in Plant and Equipment (N1; billions of current dollars)						
26. One	1.9	0.87	0.91	1.6	0.66	0.93
27. Two	3.8	1.13	0.84	2.6	0.68	0.87
28. Three	4.2	0.97	0.84	2.9	0.52	0.80
29. Four	3.8	0.70	0.91	4.2	0.56	0.82
30. Next year	2.8	0.87	0.89	2.1	0.47	0.84
Net Change in Inventories (N1; billions of current dollars)						
31. One	3.7	0.74	0.93	2.7	0.65	0.98
32. Two	4.8	1.13	0.92	3.2	1.02	1.05
33. Three	4.1	0.94	0.78	3.6	1.47	0.92
34. Four	3.8	0.70	0.71	4.3	1.31	0.85
35. Next year	3.5	0.87	0.88	2.3	0.80	1.02
Unemployment Rate (N1; per cent of labor force)						
36. One	0.2	2.09	0.68	0.1	0.92	0.92
37. Two	0.4	2.05	0.69	0.2	1.15	0.80
38. Three	0.5	2.00	0.52	0.3	1.16	0.64
39. Four	0.5	2.38	0.52	0.5	1.71	0.73
40. Next year	0.4	2.47	0.60	0.3	1.58	0.77

[a]For III-1966–III-1969. The number of forecasts included in the averages for this model, for all the variables covered and all three forecasts (*ex ante, XA; ex post, XP,* and naive, *N*) varies with the predictive span as follows: one quarter, 13; two quarters, 12; three quarters, 11; four quarters, 10; next year, 10.

[b]For II-1967–III-1969. The numbers of *XA* and *N* forecasts included in the averages for this model vary with the predictive span as follows: one quarter, 10; two quarters, 9; three quarters, 8; four quarters, 7; next year, 7. These samples of observations are used for the

Notes to Table 6 (concluded)

entries in columns 4 and 5. The number of *XP* forecasts are the same, with these exceptions: four quarters, 6; next year, 6. These somewhat smaller samples are used for the entries in column 6.

[c]*MAE-XA* denotes the mean absolute error of the *ex ante* forecasts.

[d]Ratio of *MAE-XA* to the mean absolute error of a naive model (*MAE-N*). *N*1 projections are used in the ratios for some variables and *N*2 projections for others, depending on which of the two naive models worked better in the given case. The selected model is identified in parentheses after the subtitle for each variable.

[e]Ratio of *MAE-XA* to the mean absolute error of the *ex post* forecasts (*MAE-XP*). The *ex post* forecasts use the same adjustments of constant terms that were made by the model builders in the corresponding *ex ante* forecasts. See text.

and two are larger (the poor results include the two shortest Wharton forecasts and the OBE forecasts beyond one quarter). The ratios for the predictions of the price deflator all exceed unity and are particularly large for the OBE model (lines 11–15).

The results are also poor (in the sense of "less accurate than the better of the naive models") for the shortest forecasts of spending on automobiles and for the Wharton forecasts of spending on other durables, while the other predictions for these two variables pass these tests with more success (lines 16–25). The comparisons for the two investment variables (lines 26–35) are on the whole more favorable to the forecasts, although some of the ratios, here too, are high, as, for example, the OBE predictions of the inventory change over longer spans. The worst results (highest ratios) are for the unemployment forecasts, especially in the Wharton model (lines 36–40).

It is disappointing that these forecasts, based as they are on well-known and rather elaborate econometric models, do not score better in comparison with simple naive-model projections; but this outcome must be seen in its proper historical context, and its importance must not be exaggerated. The forecast periods are short (III-1966–III-1969 for Wharton, I-1967–III-1969 for OBE) and they cover no declines in such variables as GNP in current and constant dollars and the price deflator but only strong and fairly steady upward trends. The growth trend was not much less dominant even in the expenditures for consumer durables excluding automobiles. It is not surprising that the *N*2 model gave relatively good results for these series in this particular period, and it is clear that at some other times this model would not perform nearly as well. Probably more significant is the failure of these econometric forecasts to achieve better scores relative to the *N*1 benchmark for the more volatile series: the expenditures on automobiles, investment in plant and equipment, net change in inventories, and the unemployment rate.

Table 6 also includes comparisons of *ex ante* with *ex post* forecasts in the form of ratios of the corresponding mean absolute errors, *MAE-XA/MAE-XP* (columns 3 and 6). The *ex post* forecasts incorporate the same original constant-term adjustment as the *ex ante* ones. This makes the comparisons more meaningful; moreover, the *ex post* forecasts are, on the whole, more accurate with these adjustments than with the mechanical adjustments or, a fortiori, without any adjustments at all.

Most of the ratios *MAE-XA/MAE-XP* in Table 6 are smaller than 1, particularly for the Wharton model, indicating that the *ex ante* forecasts tend to be more accurate than the *ex post* ones with the same original adjustments. However, the ratios exceed unity for the Wharton and the OBE predictions of the price deflator and of expenditures on consumer durables other than automobiles (lines 11–15 and 21–25).

The Wharton *ex post* forecasts are less accurate than the selected naive models in half of the cases examined, including most of those relating to GNP in 1958 dollars, expenditures on durables excluding automobiles, and the unemployment rate.[40] The OBE *ex post* forecasts also have larger *MAE*'s than the naive models in about half of the forty cases covered in Table 6. It is interesting to note that among the OBE predictions those of nominal GNP, expenditures on automobiles, and unemployment compare unfavorably, while virtually all others compare favorably, with the benchmark extrapolations.

3.6 The Relative Accuracy of Econometric Model Forecasts and Judgmental Forecasts

It is difficult to get meaningful comparisons of errors for these two classes of predictions. On the one hand, forecasts made without an explicitly formulated econometric model are much more numerous than forecasts made with such a model, and they also cover longer periods of time than the *ex ante* model predictions (this applies particularly to quarterly forecasts). On the other hand, for the short periods they do cover, the econometric model forecasts are generally much more regular and complete than the others. All in all, there are not very many forecasts in the two categories that are "strictly" comparable in the sense of having been made at about the same time for the same target period.

[40] Ratios of *MAE-XP* to *MAE-N* for the Wharton model can be obtained by dividing the entries in column 2 by the corresponding entries in column 3 of Table 6. "Less accurate" forecasts are those for which the ratio is greater than 1 (column 2 greater than column 3). Similarly, by taking the ratios of the corresponding entries in columns 5 and 6 of the table, *MAE-XP* can be compared with *MAE-N* for the OBE model.

Table 7 collects such forecasts for the Wharton model and four sets of business economists' forecasts and compares their mean absolute errors over spans of from one to four quarters. The forecasts refer to GNP in current and constant dollars and include both *ex ante* and *ex post* model forecasts (with the original judgmental adjustments).

The errors of the Wharton *ex ante* forecasts of GNP are on the average smaller than those of the judgmental forecasts in sets A, G, and S; in only two of these comparisons are the ratios *MAE-XA/MAE-P* larger than 1, though the ratios for set S are all high, exceeding 0.8 (Table 7, lines 1–4, 9–16). The forecasts of set C, however, tend to produce somewhat smaller errors than the Wharton model (as shown by the ratios in lines 5–8, column 5, only one of which falls below unity). But the set C forecasts themselves are in the recent years based largely on an econometric model used flexibly by the economic staff of the company that formulated it; hence these comparisons are not so much between forecasts with econometric models and judgmental forecasts as between academic and private business forecasts both of which exemplify the model-cum-judgment approach.

The Wharton *ex post* forecasts of GNP score on the whole less well relative to the judgmental forecasts. They appear to be in most cases less accurate than forecasts C and not much more accurate than forecasts G and S, while the comparisons with set A produce a very mixed picture (columns 3, 4, and 6).

For GNP in constant dollars, both the *ex ante* and the *ex post* forecasts with the Wharton model tend to have larger mean absolute errors than the predictions from sets C and S and smaller *MAE*'s than the predictions from set G (Table 7, lines 17–27).

In Table 8, which has exactly the same format as Table 7, the relative accuracy of the OBE model and the judgmental forecasts is analyzed. The *ex ante* predictions of GNP with the OBE model are better than the corresponding predictions of sets A and G (lines 1–3, 8–11, columns 2, 4, and 5). In comparisons with sets C and S, however, only the shortest OBE forecasts come out ahead (lines 4–7, 12–15).

The OBE *ex post* forecasts of GNP likewise present a mixed picture. The *MAE* ratios are here larger than 1 in most comparisons with forecasts C and S, but smaller than 1 in virtually all comparisons with sets A and G (Table 8, columns 3, 4, and 6).

Similarly to the Wharton forecasts, the OBE forecasts of real GNP, both *ex ante* and *ex post,* have for the most part larger average errors than sets C and S (but the opposite applies to some of the shortest forecasts), while set G is here definitely inferior (Table 8, lines 16–26).

TABLE 7

Comparisons of Average Errors of Wharton Model and Judgmental Forecasts over Span of from One to Four Quarters, 1966–69
(GNP is in current dollars in lines 1–16; in constant dollars in lines 17–27)

Span of Forecast (no. of quarters)	No. of Forecasts in Each Average[a] (1)	Mean Absolute Errors (MAE) for the: *Ex Ante* Model Forecasts (*MAE-XA*) (2)	*Ex Post* Model Forecasts (*MAE-XP*) (3)	Forecasts in Sets A. . .S (*MAE-P*)[b] (4)	*MAE* Ratios: *MAE-XA*/*MAE-P*[c] (5)	*MAE-XP*/*MAE-P*[d] (6)
		Set A (I-1967–III-1969)				
1. One	3	5.1	12.5	6.6	0.77	1.90
2. Two	3	6.9	10.0	11.9	0.58	0.84
3. Three	3	8.8	5.3	13.5	0.65	0.39
4. Four	2	14.8	14.4	14.0	1.06	1.03
		Set C (III-1966–III-1969)				
5. One	6	4.8	8.1	3.4	1.41	2.38
6. Two	6	7.1	8.4	7.0	1.01	1.20
7. Three	6	9.8	9.5	10.9	0.90	0.87
8. Four	5	15.2	15.8	11.8	1.29	1.34
		Set G (I-1967–III-1969)				
9. One	6	5.2	8.4	7.5	0.69	1.12
10. Two	5	8.0	8.3	15.2	0.53	0.84
11. Three	5	10.3	7.1	21.2	0.49	0.87
12. Four	4	17.1	12.6	29.8	0.57	0.94
		Set S (III-1968–III-1969)				
13. One	5	4.9	5.6	5.7	0.86	0.98
14. Two	4	10.5	8.9	10.6	0.99	0.84
15. Three	3	18.2	15.0	17.2	1.06	0.87
16. Four	2	19.1	22.6	24.0	0.80	0.94
		Set C (III-1966–III-1969)				
17. One	6	4.2	6.8	2.8	1.50	2.43
18. Two	6	5.6	6.9	4.9	1.14	1.41
19. Three	6	6.3	9.2	6.0	1.05	1.53
20. Four	5	7.4	14.2	4.8	1.54	2.96

Table 7 (concluded)

			Set G (I-1967–III-1969)			
21. One	6	4.1	7.1	5.6	0.73	1.27
22. Two	5	5.9	6.2	11.2	0.53	0.55
23. Three	5	5.6	6.2	16.0	0.35	0.49
24. Four	4	6.6	10.0	22.5	0.29	0.44
			Set S (III-1968–III-1969)[e]			
25. One	4	1.9	4.7	1.9	1.00	2.47
26. Two	3	3.8	6.3	1.1	3.45	5.73
27. Three	2	4.9	7.6	0.8	6.12	9.50

[a]Refers to the corresponding entries in each of columns 2, 3, and 4.

[b]*P* denotes predictions in one of the sets identified in the subtitles below (set A, C, G, or S). Set S represents the American Statistical Association–NBER quarterly Business Outlook Survey (group median forecasts). For a brief description of the other forecast sets, see text.

[c]Ratio of column 2 to column 4.

[d]Ratio of column 3 to column 4.

[e]The four-quarter forecasts could be compared for a single period only and are therefore omitted.

The *MAE* ratios for either model show no general and systematic tendency either to increase or decrease with the span of the forecast. However, the ratios do increase markedly with the span in some cases, notably for the OBE *ex ante* forecasts in comparisons with set C and all the model forecasts in comparisons with set S (Table 7, lines 25–27, and Table 8, lines 16–19 and 24–26). Thus, possibly with the above exceptions, these groups of forecasts apparently do not possess any strong differential advantages or disadvantages that would depend specifically on the span of forecast.

To sum up these findings, the *ex ante* forecasts with Wharton and OBE models hold a slight edge over the judgmental forecasts, their average errors being the smaller ones in 58.5 per cent of the cases.[41] The *ex post* forecasts with these models are on balance only about as accurate as the forecasts in the other four sets (the two models show smaller errors in 50.9 per cent of the cases).[42]

These results should not be used for any extensive generalizations and must indeed be interpreted very cautiously because they are based on small samples limited (as already noted) by the scarcity of data for different

[41]For the Wharton model alone, the percentage is 55.6; for the OBE model, 61.5.

[42]Again, to quote the corresponding percentages separately for the two models: Wharton, 48.1 per cent; OBE, 53.8 per cent.

TABLE 8

Comparisons of Average Errors of OBE Model and Judgmental Forecasts over Spans of from One to Four Quarters, 1967–69

(GNP is in current dollars in lines 1–15; in constant dollars in lines 16–26)

Span of Forecast (no. of quarters)	No. of Forecasts in Each Average[a] (1)	Mean Absolute Errors (MAE) for the *Ex Ante* Model Forecasts (MAE-XA) (2)	*Ex Post* Model Forecasts (MAE-XP) (3)	Forecasts in Sets A. . .S (MAE-P)[b] (4)	MAE Ratios: MAE-XA[c] / MAE-P (5)	MAE-XP[d] / MAE-P (6)
		Set A (I-1968–III-1969)[e]				
1. One	2	1.8	6.9	6.8	0.26	1.01
2. Two	2	11.0	2.7	13.4	0.82	0.20
3. Three	2	14.8	4.9	17.8	0.83	0.28
		Set C (III-1967–III-1969)				
4. One	4	2.9	7.1	3.9	0.75	1.82
5. Two	4	10.6	9.4	7.1	1.49	1.32
6. Three	4	15.5	13.7	14.6	1.06	0.94
7. Four	3	21.3	19.9	18.4	1.18	1.08
		Set G (III-1967–III-1969)				
8. One	5	2.4	5.8	8.7	0.28	0.67
9. Two	4	10.6	9.4	17.0	0.62	0.55
10. Three	4	15.5	13.7	21.8	0.71	0.63
11. Four	3	21.3	19.9	31.7	0.67	0.63
		Set S (III-1968–III-1969)				
12. One	5	2.9	4.8	5.6	0.52	0.86
13. Two	4	9.7	12.7	10.6	0.92	1.20
14. Three	3	17.3	21.2	17.1	1.01	1.24
15. Four	2	24.2	30.6	23.7	1.02	1.29
		Set C (III-1967–III-1969)				
16. One	4	2.3	4.2	3.4	0.68	1.24
17. Two	4	6.5	4.3	5.5	1.18	0.78
18. Three	4	9.8	7.5	7.4	1.32	1.01
19. Four	3	12.8	11.6	5.7	2.25	2.04

Table 8 (concluded)

			Set G (III-1967–III-1969)			
20. One	5	2.0	3.7	6.5	0.31	0.57
21. Two	4	6.5	4.3	13.3	0.49	0.32
22. Three	4	9.8	7.5	17.4	0.56	0.43
23. Four	3	12.8	11.6	25.5	0.50	0.45
			Set S (III-1968–III-1969)[e]			
24. One	4	1.3	1.5	1.9	0.68	0.79
25. Two	3	2.0	3.0	1.2	1.67	2.50
26. Three	2.	2.5	3.0	1.2	2.08	2.50

[a]Refers to the corresponding entries in each of the columns 2, 3, and 4.

[b]*P* denotes predictions in one of the sets identified in the subtitles below (set A, C, G, or S). Set S represents the American Statistical Association–NBER quarterly Business Outlook Survey (group median forecasts). For a brief description of the other forecast sets, see text.

[c]Ratio of column 2 to column 4.

[d]Ratio of column 3 to column 4.

[e]The four-quarter forecasts could be compared for a single period only and are therefore omitted.

forecasts that can be matched by span and target period. Since intertemporal variations in errors of predictions from a given source are often large, it is possible that at least some of the details in the picture presented by these comparisons reflect strongly the particularities of the periods covered and would not be upheld by the evidence from larger samples.[43]

4. Suggested Agenda for Future Research

4.1 Focus and Feasibility of the New Studies

The results of past and current studies summarized in the two preceding sections of this report, and the experience and materials accumulated in the

[43]Comparisons of the *MAE-XA* figures in Table 6, columns 1 and 4, with those in Tables 7 and 8, columns 2, show that the average errors in GNP forecasts with both the Wharton and the OBE model have been for the most part larger in the selected periods than in the total forecast periods. Of the other GNP forecasts, set C appears to be favored by the selection, in the sense of having a smaller *MAE* in the periods covered in Tables 7 and 8 than at other times, whereas the opposite applies to set G. If average errors of all GNP predictions available from each of these sources were used, that is, if all differences in the periods covered were disregarded, the comparisons with some of the other forecasts, notably those of set C, would be definitely more favorable to the *ex ante* forecasts with the two econometric models. For the *ex post* forecasts and the real GNP predictions, the results of a similar analysis are mixed.

process of this research, indicate both the need for and the promise of further work on short-term economic forecasts and forecasting models. They also provide a good basis for planning and executing such work.

The planned research consists of several related projects that share a focus on the problem of how predictive accuracy is related to the method and structure of the forecasts. This presupposes (a) that the accuracy of predictions of different types and from different sources can be measured and compared, and (b) that the methods used by forecasters and the ingredients of forecasts can be ascertained and analyzed. No doubt, neither of these conditions is fully met, but, here as elsewhere, it is necessary to do the best one can with unavoidably incomplete information and imperfect knowledge. I am prepared to argue that, as a consequence of previous and current work, it is now possible to undertake studies so oriented with a fair chance of making a significant contribution.

Thus, regarding (a), recent studies (including those reviewed in section 2) have achieved considerable progress in developing criteria of forecasting accuracy and applying them to diverse forecasting data. There is much less success to report on matters relating to (b), but it has been shown that useful information about "judgmental" forecasts (defined as those that are not based on an explicit forecasting "model") can be gained from comparisons of such forecasts with predictions derived by a known, replicable method.[44] Moreover, direct information is now being systematically collected about the forecasting methods used by the participants in the new ASA-NBER Quarterly Survey of the Economic Outlook.[45] These data, then, permit a cross-classification by both accuracy and method of a sample of business forecasts for any particular period surveyed and over time. When sufficient material from the successive surveys becomes available, it should be possible to learn from it whether the quality of forecasts varies systematically with the method used, and if so how. In seeking an answer to this question, one will be able to control for several factors: the date of issue of the forecasts (the survey questionnaires are collected at certain fixed times), the recent and base-period estimates (they are given or reported), and the key assumptions about tax and monetary policy, defense posture, government spending, and so on (the forecasters are asked to specify them and most do).

The econometric forecasts, of course, do rely on explicitly formulated models and certain more or less standardized methods of statistical estimation. However, the *ex ante* forecasts with econometric models must use judgmental predictions or estimates of the exogenous inputs, and they

[44] See [16], [17], and [29], *passim*.

[45] See [32] and [33].

typically use various adjustments of the constant terms, which likewise involve judgment and cannot as a rule be reproduced by extrapolative or other mechanical procedures.[46] In the past, such operations were frequently not well recorded, and experience shows that it is difficult and sometimes impossible to reconstruct econometricians' forecasts without this vital documentation. It is therefore indispensable for the studies here contemplated that a systematic collection of forecasts from several selected econometric models be carried out in such a way as to include not only the predictions that were made but also the judgmental inputs and adjustments which, along with the known structure of the model, produced these predictions. Full information of this type is necessary for independent replication of the forecasts, to ensure that the appropriate models and data (inputs and outputs) are being recorded. Steps have already been taken at the National Bureau to initiate such a collection of quarterly econometric forecasts and related statistics. The Wharton-EFU and the Commerce-OBE models are being processed, as is the new quarterly Michigan model, and it is hoped that other models will be added. Of course, the cooperation of the model builders is essential for this effort and its success.[47]

4.2 Comparisons of Types and Methods of Forecasting

Economic forecasts can be classified by several different criteria such as (a) the degree to which they use formalized methods vis-à-vis informal judgment; (b) the degree to which they rely on relationships among different variables vis-à-vis projections of past behavior of the variable to be predicted; and (c) the degree to which they are based on averages of macropredictions or aggregations of micropredictions from different sources vis-à-vis global forecasts from a single source. According to (a), a whole gradation could be established, from forecasts that are "purely judgmental" to those produced by fully specified and strictly implemented econometric models. According to (b), similarly, there is, at least in principle, an ordering of forecasts from pure extrapolations of past values of the given variable (inferring X_{t+i} from X_{t-j}, where i and j denote the various discrete or distributed lags that may be involved) to outputs of equations in which X_{t+i} is related to other variables only, say Y_{t-j}. According to (c), finally, forecasts would be divided into predictions by individuals and small teams, weighted or unweighted averages

[46]For a description and evaluation of such adjustments, see [9, section III and IV] and [12, section 3.1].

[47]This may involve some restrictions on the scope or timing of the release of the resulting information, but arrangements of this sort have already been made in some cases and they should not pose any critical problems.

of such predictions (including small group forecasts and large opinion polls), aggregates based on surveys of businessmen's or consumers' anticipations, and so on.

These various classes of forecasts overlap and can be combined in diverse ways; the pure or extreme forms are generally used only as benchmarks of predictive preformance, not as forecasts proper. For example, a forecast of GNP and its major components may combine extrapolation, relationship of the target series to known or estimated values of some other variables, external information such as a survey of investment intentions and government budget estimates, and the judgment of the forecaster. This can be and often is true of a "judgmental" forecast by a business economist as well as of a "model" forecast by an econometrician. The main difference between the two would be that the econometrician's model is explicitly formulated in quantitative terms, whereas the business economist's is not. Mechanical methods of extrapolating the series to be predicted and of relating it to other series would then serve to provide standards against which to measure the performance of the bona fide forecasts made by identifiable sources prior to the event. Where such forecasts relate to variables covered by the surveys of anticipations or intentions, the aggregated results of these surveys furnish another important yardstick of predictive accuracy.

The currently relevant forecasting procedures can be broadly classified as (a) extrapolative techniques; (b) surveys of intentions or anticipations by economic decision-making units; (c) business cycle indicators; and (d) econometric models. Each of the four methods furnishes data that constitute ingredients of actual forecasts of the course of the economy; and each of these four classes of data, too, can be used to construct mechanically reproducible predictions that represent appropriate benchmarks for forecast evaluation. In large part, the proposed research strategy consists in confronting forecasts of different types with an array of these benchmark predictions in such a way as to extract from the comparisons as much information as possible about both the accuracy and structure of the forecasts.

To use for illustration a case particularly rich in opportunities for such comparisons, consider an econometric model's equation for business expenditures on plant and equipment. Forecasts of this variable produced by the econometrician armed with this model can be confronted by: (a) corresponding extrapolations from different "naive models," including some relatively sophisticated and effective types of autoregression; (b) aggregated business anticipations of plant and equipment outlays (based, say, on the OBE-SEC surveys); (c) forecasts from regressions including some appropriate "leading indicators" such as new orders and contracts for plant and equipment or new capital appropriations; and (d) *ex post* forecasts based on the model as a whole, or alternatively, on the single equation concerned, with actual values used for the exogenous inputs and no judgmental adjustments of the constant terms.[48] The model may, of course, incorporate one or more of these elements in some form and may also have other ingredients of net predictive value, in which case it should prove superior to the benchmark(s) in question; e.g., if the investment equation includes a lagged value of the dependent variable along with other significant explanatory factors, then the forecasts based on it will be better than the standard offered by predictions from the corresponding autoregressive model. But such an outcome is by no means always to be expected, for the model may not use fully or optimally the information embodied in the extrapolative, anticipatory, or indicator benchmark models, and it may contain other misspecifications. Where the comparisons indicate such weaknesses, it is a logical next step to try to improve the model by adding to it the elements that seem promising and removing those that seem disturbing in the light of these tests.

Such studies, then, should not only contribute to our knowledge of the relative accuracy of different forecasts and forecasting models (including the benchmark models), but they should also help locate some of the areas of strength and weakness in the underlying methods and specifications. An analysis of this kind is likely to yield the highest returns for the forecasts based on explicit models, but even where such models are lacking the payoff may be considerable. For example, suppose one can demonstrate that business forecasts from a particular source would have gained in the past by being combined with extrapolations of some specific type (say, by means of weights derived from a regression of the actual values on both the forecasts and the extrapolations). If this is a sufficiently stable relationship, it may be indicative of a systematic and correctable deficiency of the forecast.

Clearly, there are many examples in the literature of such comparisons as are here proposed, and some useful lessons can be drawn from them. The advantage of the suggested approach is that it combines these familiar elements into a framework for a more comprehensive and systematic analysis of economic forecasting than has been hitherto attempted.

[48]Alternatively, certain mechanical adjustments could also be employed. In systems with lagged terms of the endogenous variables, model-generated values for these factors must be used to generate multiperiod forecasts (in single-equation models, such inputs are of course limited to the values of one dependent variable). The use of anticipatory data tends to limit the time span over which a model can be successfully extrapolated into the future because such data are themselves difficult to predict well, as noted at the end of section 2.8.

4.3 The Qualities of Short-Term Econometric Model Forecasts

In continuation of new and significant research undertakings in this area,[49] *ex ante* and *ex post* forecasts of the principal endogenous variables need to be compared for each of the major quarterly models of the U.S. economy listed at the end of section 4.1. It is necessary to have several different variants of this analysis: with and without such judgmental adjustments as were made by the model builders for their forecasts, and also with the most promising of the mechanical adjustments of the constant terms. The results will permit assessments of (a) the magnitude and effects of errors in the projections of the exogenous variables; (b) the quantitative impact and relative efficiency of the different types of adjustments; and (c) the magnitude of forecast inaccuracies that are attributable to errors in construction and solution of the given model. It will then also be possible to make certain inferences about the interaction of errors from these different sources.

Both *ex ante* and *ex post* model forecasts should then be confronted by various "naive" and autoregressive extrapolations, a sample of judgmental forecasts, single-equation predictions, and (if appropriate) aggregated micro-anticipations data from surveys. Efforts should be made to standardize at least the sample period used, so that meaningful intermodel comparisons can be made of the accuracy and other properties of the forecasts.

This analysis is expected to help provide answers to several interesting questions about the relative merits and demerits of different models and forecasts. To list some that seem particularly important: (a) What are the desirable properties of measures of goodness of predictions? (b) Are predictions from econometric systems superior to noneconometric predictions? (c) Do large-scale models perform better than small-scale models in forecasting GNP and its major components? (d) Insofar as they yield to our measurement efforts, how do the benefits and costs of econometric-system forecasts compare with those of single-equation predictions? (e) With what success for future prediction can technical adjustments be made on the basis of the prediction record?

Some of these questions, such as (b) and (e), have already received partial and tentative answers in [9] and in section 3 of this report, but the analysis needs to be extended to materials for other model and nonmodel forecasts and to the remaining and related questions. The most basic and penetrating ones among the latter deal with misspecification errors in the models covered: (f) Where do the most serious errors of this kind originate in each of the

[49] See references to [9] and to related ongoing research mentioned in earlier parts of this report.

models? (g) How do they affect the forecasting qualities of the models? (h) What changes in the models would be likely to reduce such errors and improve the forecasts? (i) What are the properties of a sound short-term predictive model——do they differ from those of a good structural model and, if so, how?

Even a very comprehensive evaluation of the forecasting aspects of the models will not be sufficient to settle questions of this type, but some contributions to the answers should be forthcoming from this work when combined with the simulation analyses that are discussed below. However, a more effective treatment of misspecification problems presumably requires a thorough review of a model, by individual structural equations and by blocks of closely related equations for a given economic sector or process. Re-estimation of a given block, with the variables in other blocks being taken as exogenous, would be desirable, so as to detect those results that are in conflict with theoretical expectations (the latter refer generally to such conditional predictions). This examination of block simulations and predictions could at least help isolate those instances in which the signs of the model-generated statistics appear objectionable; the theory has less to say about the magnitudes of the various coefficients, multipliers, and elasticities involved, but the procedure should be able to spotlight any estimates that are grossly out of line. In this way, one might hope to trace the misspecifications to the individual equations, which should then be revised; after which the "corrected" model should be subjected to new exercises in simulation and prediction designed to test its performance. There is a very wide range of variation on this general theme, and a number of approaches appear promising; e.g., separate estimates for periods of cyclical expansions and contractions could be computed and examined in an effort to test the effects of nonlinearities in the models. The need for all such tests has been emphasized, and there are some notable initiatives in this direction in recent literature.[50] However, these are limited critical articles by individuals, and what is required here is clearly a comprehensive and systematic study by a group of researchers.[51]

[50]Two papers on the Brookings model should be noted in this context: Zvi Griliches, "The Brookings Model Volume: A Review Article," *Review of Economics and Statistics*, May 1968, pp. 215–234 (see also "Comment" on this article by Gary Fromm and Lawrence R. Klein, *ibid.*, pp. 235–240); and Robert J. Gordon, "The Brookings Model in Action: A Review Article," *Journal of Political Economy*, May–June 1970, pp. 489–525.

[51]It would of course be highly desirable for such a project to secure the cooperation of the model builders, but the latter should preferably not be included in the research group itself.

4.4 Business Cycle Analysis of Econometric Model Simulations

Simulation studies of the quarterly U.S. models can help answer some of the questions formulated above (a, c, e, f, g, h, and i) and others such as these: (j) Do the models under review generate cyclical behavior as defined and observed in the empirical business cycle studies, notably those of the NBER? (k) If so, to what extent are such fluctuations in the estimated series produced endogenously by the models, and to what extent are they attributable to external impulses? In the process, much can be learned from the models about the qualities of the *ex post* or conditional predictions, e.g., about the sensitivity of the simulations to the initial conditions and the length of the predictive span. Such findings are likely to lead to inferences about the structural properties of the models; thus, a tendency for nonstochastic simulations to produce strongly *damped* fluctuations in projected GNP would suggest that the model in question is dynamically stable. Last but not least, insofar as the models provide valid representations of the economy, simulation studies hold out a promise of contributing to the knowledge of business cycles "in the real world," the question here being, Do the cycles consist mainly of endogenous or exogenous movements?

As an outgrowth of the 1969 Harvard conference and owing to the cooperation of the builders of several models, there now exists a very large volume of rich statistical materials bearing on these problems. A cyclical analysis of some of these data was undertaken in [34]. This study was seriously limited by the resources available, including time, but its results met with considerable interest and a consensus that further work in this area was definitely desirable.[52]

To produce any cyclical movements, the models examined so far apparently require perturbations in either the exogenous variables or the relationships with endogenous variables or both. In experimenting with 100-quarter stochastic simulations that start at very recent dates and extend into the future, smooth trend projections for the exogenous variables were used in [34], but the absence of shocks or fluctuations in these projections is an unrealistic feature which could be at least partly responsible for the weakness of the cyclical elements in the simulations analyzed. Further tests are needed to determine whether this weakness can be remedied or reduced by imposing more or less sporadic disturbances on the exogenous factors.

Only small samples of the stochastic simulations have as yet been studied for most of the endogenous variables selected for this investigation. The

[52]This conclusion is based on the discussions at the Conference on Econometric Models of Cyclical Behavior cosponsored by the Social Science Research Council and the National Bureau of Economic Research, Cambridge, Mass., November 14, 1969, and at the follow-up meeting to that conference at the NBER, New York, February 27, 1970.

analysis needs to be extended to larger samples and to models other than those included in [34].

4.5 Interaction of Economic Developments and Policy Changes

Another important use of simulation techniques is for the study of the effects upon the economy of alternative changes in the exogenous policy variables. All full-scale models have variables of this kind, which are specified as being determined by the actions of fiscal or monetary authorities. However, the choice of these policy factors varies between the different models, and some factors, particularly in the monetary area, have only recently begun to receive explicit attention; the resulting discrepancies impair intermodel comparisons of policy sensitivity.

Moreover, there are feedback effects of developments in the economy upon the policy variables, which are very much in need of being explored. For example, Federal Reserve policy is not necessarily exogenous in the sense that it cannot be explained by other parts of the system; that policy might be assumed to respond to the behavior of prices or unemployment or the balance of payments (or to all of these and to still other factors), and is exogenous only in the sense that the monetary authorities are not constrained to take specified actions because of competitive or other kinds of market forces. A systematic relation capable of being described in the model could exist between, say, the changes in national output and the price level and the reaction of the monetary and fiscal authorities as reflected by movements in the so-called exogenous policy variables.

In a paper prepared for this colloquium, Yoel Haitovsky and Neil Wallace discuss the results of experimental work with stochastic simulations designed to reveal the effects of such interactions. They impose upon a model some alternative fixed "policy rules," which involve certain quantified monetary and fiscal responses to changes in the unemployment rate and the implicit price deflator for GNP. The model includes autoregressive equations for the noninstrumental exogenous variables, and the parameters and additive disturbances are chosen randomly for each run or from quarter to quarter. The experience gained in this pilot study should prove highly valuable for future research in this area, which would probably involve extensions of the analysis to a variety of models, policy rules, and instruments.

4.6 Feedback Effects of Major Macroeconomic Forecasts

A closely related problem concerns the feedback effect attributable to publication of the forecast. Certain influential forecasts may significantly affect the behavior of economic agents; the resulting changes in the economy might in turn lead to revisions of the policies, which would alter the general economic development, and so on.

Published official forecasts (such as those of the Council of Economic Advisers in the *Economic Report of the President*) are of particular interest here, since they may well carry special weight as expressions or instruments of the government's economic policies. The accuracy of such forecasts should be compared not only with the accuracy of reputable private forecasts but also with that of government agency forecasts that are made ahead of the event but are not published or released until some later time. Both the relative accuracy of the official forecasts and their influence on private forecasts and expectations are matters of very considerable interest. There is also need to study government forecasts in the context of concurrent and subsequent economic policies, so as to learn more about how these factors interact.

In some countries (Holland, Sweden, France), highly authoritative and influential forecasts, based to a substantial extent on econometric models, are prepared by agencies of governments that assume very active roles in macroeconomic planning and policies. (In the United States, there is much less centralization in these affairs.) In such situations, publication of the government forecast may be interpreted to mean that steps will be taken to change policy variables in the event that private actions are inconsistent with the forecast results. In effect, under some circumstances the forecast might be more accurately described as a plan to be implemented. Private reactions are likely to be different if the forecast is viewed as a plan that can influence policies rather than if no such connection between prediction and the policy-making power is perceived. For these reasons, the feedback effects of forecasts would probably lend themselves best to study by means of selected international comparisons of forecasts and policies.

How effective are forecasts in influencing policies and general expectations, and thereby their own accuracy? Are forecasts frequently and significantly validated when the predicted developments appear desirable to those whose decisions count and are they, analogously, invalidated when they predict changes that are deemed undesirable? Of the interesting but limited theoretical work on such problems, some is highly abstract and speculative, and little is conclusive.[53] Yet, with the presently available models and data,

[53]The main contributions are: Emile Grunberg and Franco Modigliani, "The Predictability of Social Events," *Journal of Political Economy*, December 1954, pp. 465–478; Murray Kemp, "Economic Forecasting When the Subject of the Forecast Is Influenced by the Forecast," *American Economic Review*, June 1962, pp. 492-496; and the Communications on the latter paper by A. C. Chiang, Grunberg and Modigliani, and Kemp, *American Economic Review*, September 1963, pp. 730–740.

better results should be obtainable, and considerable interest apparently exists in international comparisons of economic forecasts generally and in the specific questions raised here.[54]

4.7 Other Proposals and Concluding Remarks

Other related projects in this general area of research are being planned at the National Bureau, and some of them appear to be as important and as promising as those discussed above; but this report is long already and must be kept within reasonable bounds, so only brief references can be made to these proposals. One of the potentially most useful undertakings, as far as improvement of the short-term predictive accuracy of econometric models is concerned, is to revise the models in the direction of making more and better use of survey data on business and consumer anticipations, governmental budget estimates, leading indicators, and other similarly forward-looking information. Besides enhancing the forecasting ability of the models, this work may also yield returns in increased understanding of the economic processes involved and in the feedback on the surveys, which could suggest ways of obtaining more useful expectational data.

Another newly opened field of study that deserves much attention would combine estimation of monthly aggregative series, where only quarterly data now exist, with construction and examination of a monthly econometric forecasting model of the U.S. economy. The feasibility of constructing a monthly recursive macromodel was recently tested by T. C. Liu, with generally positive results.[55] Monthly series for GNP and its major components can be estimated indirectly by the method of interpolation by related series for which monthly observations are available, although much remains to be done to establish how best to implement this approach. Once a monthly model, similar in size, structure, and estimation procedure to an existing quarterly model, is constructed, the relative predictive performance of the two could be compared by means of an analysis of corresponding *ex post* forecasts and simulations. The dynamic properties of the monthly model could be examined by spectral analysis and by business cycle analysis of model simulations. This research, too, could provide some insights into the consequences of aggregating an econometric model over time. It would

[54]This opinion is based on exploratory discussions that the author had with several prospective suppliers of statistical data and other information pertinent to such a study: the directors or officers in charge of forecasting at the leading economic research institutes and planning agencies in London, Stockholm, Paris, and the OECD.

[55]See Ta-Chung Liu, "A Monthly Recursive Econometric Model of United States: A Test of Feasibility," *Review of Economics and Statistics*, February 1969, pp. 1–13.

probably be sufficient to use small-scale models, at least in the pilot studies contemplated here.

In the nature of things, *plans* for future research permit no *conclusions*, so only a few general remarks are offered to conclude this report. The investment in research resources embodied in the past and current National Bureau studies on economic forecasting is believed to have paid off well in the results which are partly and briefly summarized in sections 2 and 3 above; but the verdict on this must be left to the reader and the user of these studies and, ultimately, to those who will judge their influence on and usefulness for future scientific inquiry. The last part of my paper will have accomplished much of its purpose if it showed how much remains still to be done, while also conveying the sense that ambitious efforts in this field are now both needed and promising.

REFERENCES

1. Christ, Carl F. "A Test of an Econometric Model of the United States, 1921–1947." In [26, pp. 35 – 107].
2. ––– . "Aggregative Econometric Models: A Review Article," *American Economic Review*, June 1956, pp. 385–408.
3. Cole, Rosanne. "Data Errors and Forecasting Accuracy." In [15, pp. 47–82].
4. ––– . *Errors in Provisional Estimates of Gross National Product.* New York, National Bureau of Economic Research, 1969.
5. Conference on Research in Income and Wealth. *Short-term Economic Forecasting.* Studies in Income and Wealth, Vol. 17. Princeton University Press for NBER, 1955.
6. Cox, Garfield V. *An Appraisal of American Business Forecasts.* Rev. ed. Chicago, University of Chicago Press, 1930.
7. Cunnyngham, Jon. "Econometric Model Forecasts." *Forty-fifth Annual Report of the National Bureau of Economic Research.* New York, June 1965, pp. 60–64.
8. Diller, Stanley. "Expectations in the Term Structure of Interest Rates." In [15, pp. 112–116].
9. Evans, Michael K.; Haitovsky, Yoel; and Treyz, George I.; assisted by Vincent Su. "An Analysis of the Forecasting Properties of U.S. Econometric Models," In *Econometric Models of Cyclical Behavior.* Studies in Income and Wealth, Vol. 36, New York, NBER, 1972.
10. Evans, Michael K., and Klein, Lawrence R. *The Wharton Econometric Model.* Studies in Quantitative Economics No. 2. Philadelphia, Economics Research Unit, University of Pennsylvania, 1967.
11. Fels, Rendigs, and Hinshaw, C. Elton. *Forecasting and Recognizing Business Cycle Turning Points.* New York, NBER, 1968.
12. Green, George R., in association with Maurice Liebenberg, and Albert A. Hirsch. "Short- and Long-Term Simulations with the OBE Econometric Model." In *Econometric Models,* (see [9]).
13. Juster, F. Thomas. "Consumer Anticipations and Models of Durable Goods Demand." In [15, pp. 167–242].

REFERENCES (continued)

14. Liebenberg, Maurice; Hirsch, Albert A.; and Popkin, Joel. "A Quarterly Econometric Model of the United States: A Progress Report." *Survey of Current Business,* May 1966, pp. 13–39.
15. Mincer, Jacob, ed. *Economic Forecasts and Expectations: Analyses of Forecasting Behavior and Performance.* New York, NBER, 1969.
16. Mincer, Jacob, and Zarnowitz, Victor. "The Evaluation of Economic Forecasts." In [15, pp. 3–46].
17. Mincer, Jacob. "Models of Adaptive Forecasting." In [15, pp. 83–111].
18. Moore, Geoffrey H., and Shiskin, Julius. *Indicators of Business Expansions and Contractions.* New York, NBER, 1967.
19. --- . "Forecasting Short-Term Economic Change," *Journal of the American Statistical Association,* March 1969, pp. 1–22.
20. Okun, Arthur M. "A Review of Some Economic Forecasts for 1955–1957." *Journal of Business,* July 1959, pp. 199–211.
21. --- . "On the Appraisal of Cyclical Turning-point Predictions." *Journal of Business,* April 1960, pp. 101–120.
22. Rasche, Robert H., and Shapiro, Harold T. "The FRB-MIT Econometric Model: Its Special Features." *American Economic Review,* May 1968, pp. 123–149.
23. Stekler, Herman O. *Economic Forecasting.* New York; Praeger, 1970.
24. Suits, Daniel B. "Forecasting and Analysis with an Econometric Model." *American Economic Review, March 1962,* pp. 104–132.
25. Theil, Henri. *Economic Forecasts and Policy.* 2nd ed. Amsterdam, North-Holland, 1961.
26. ---- . Applied Economic Forecasting. *Chicago, Rand-McNally, 1966.*
27. Universities–National Bureau Committee for Economic Research. *Conference on Business Cycles.* New York, NBER, 1951.
28. --- . The Quality and Economic Significance of Anticipations Data. *Princeton University Press for NBER, 1960.*
29. Zarnowitz, Victor. *An Appraisal of Short-Term Economic Forecasts.* New York, NBER, 1967.
30. --- . "Eine Auswertung von Kurzfristigen Wirtschaftsvoraussagen in den USA." *IFO-Studien,* 1966, pp. 1–41.
31. --- . "An Evaluation of Price-Level Forecasts." *1968 Proceedings of the Business and Economic Statistics Section.* American Statistical Association. Washington, D.C., 1968, pp. 284–289.
32. --- . "The New ASA-NBER Survey of Forecasts by Economic Statisticians." *The American Statistician,* Februrary 1969, pp. 12–16.
33. --- . "The ASA-NBER Quarterly Survey of the Economic Outlook: An Early Appraisal." *1969 Proceedings of the Business and Economic Statistics Section.* American Statistical Association. Washington, D.C., 1969, pp. 241–249.
34. Zarnowitz, Victor; Boschan, Charlotte; and Moore, Geoffrey H.; with the assistance of Josephine Su. "Business-Cycle Analysis of Econometric Model Simulations." In *Econometric Models* (see [9]).