Volume Title: An Appraisal of Short-Term Economic Forecasts

Volume Author/Editor: Victor Zarnowitz

Volume Publisher: UMI

Volume ISBN: 0-87014-445-6

Volume URL: http://www.nber.org/books/zarn67-1

Publication Date: 1967

Chapter Title: YARDSTICKS OF PREDICTIVE PERFORMANCE

Chapter Author: Victor Zarnowitz

Chapter URL: http://www.nber.org/chapters/c0941

Chapter pages in book: (p. 83 - 120)

6

YARDSTICKS OF PREDICTIVE PERFORMANCE

EXTRAPOLATIVE MODELS

The average error measures show how much the forecasts deviate from the unattainable ideal of no error. More realistic criteria are necessary to take account of the properties of the variables being forecast and the degree of difficulty they present to the forecaster.

The past record of the series to be predicted contains information that may have predictive value and should be exploited by the forecaster. To make use of it, some mechanical procedure is needed to extrapolate the past; since no other variables are employed, there is no place here for a theory of what economic factors determine the variable being predicted or how they operate. In this sense of being innocent of economic theory proper, all extrapolations are "naive" models (as they have been called) including some that are technically rather sophisticated.

The simplest extrapolative models, which have long been in use for this purpose, are indeed naive. The first of these (N1) predicts that the next period's value of a given variable will equal this period's value; another (N2) projects forward the last recorded change in a series, thus replacing the "same level" assumption of the first model with a "same change" assumption.[1]

Clearly, naive models of this kind should be regarded merely as providing minimal standards, which may be useful for screening out poor forecasts but are not sufficient to rate the surviving forecasts as neces-

[1] N1 can be said to specify that $A_{t+1} = A_t + u_{t+1}$, where 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 $P_{N2} = A_t^* + \Delta A_t^*$.

sarily good. The ideal extrapolation would be the one which takes best advantage of the historical content of the series for the predictive purpose, that is, the one that in the long run yields smaller errors than any other extrapolation. Then, a comparison of a forecast proper with such an extrapolation would show whether the forecaster's judgment or knowledge of the relevant economic relations gives better results than those obtained through the optimum use of the predicted variable's own record.[2]

To achieve the optimum extrapolation, a sufficiently long and consistent time series record for a given variable and a knowledge of its structure are necessary. At best, it might be possible to move closer to that ideal goal by constructing improved extrapolative models with the aid of trend projections, autoregressive schemes, and other devices.[3]

In one such approach, we use a simple but frequently rather effective type of extrapolation, which adds the average value of past changes in the series to the latest level of the series. This model, called N2*, is a projection of the mean historical change, an extension of model N2, which projects merely the last recorded change. For series with pronounced trends which are approximately linear over the periods covered, model N2* has a considerable advantage.[4]

Another approach uses an autoregressive model (N3), in which the present level is taken as a linear function of the preceding levels of the series, plus a random error.[5] This method imposes greater requirements on the data than the other, simpler models. The relationships between present and past values cannot be measured with reasonable confidence for any single series as short as those used here. Only very simple relations can be detected within such series. Obtaining large

[2] Jacob Mincer developed a method of comparing a forecast with an extrapolation which in effect decomposes them both into their common elements and their separate predictive contributions. A paper on the criteria of forecast evaluation by Mincer and myself, "The Evaluation of Economic Forecasts," will include a description and illustration of that method.

[3] Work in this area is continuing. See the reports by Jon Cunnyngham in the National Bureau Annual Reports for 1964 and 1965.

[4] 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. Hence, the forecast $P_{N2}^* = A_t^* + \overline{\Delta A}$.

[5] The general form of N3 is $A_{t+1} = a + b_1A_t + b_2A_{t-1} + b_3A_{t-2} + \dots u_{t+1}$. The forecast uses estimates of a and b's from regressions based on the values of A as available in the base period t (these are typically preliminary, A_1^* at least for t). The value u_{t+1} is assumed zero.

numbers of usable observations for the variables concerned is, however, also very difficult.[6]

The extrapolative models have been developed in quarterly terms in order to appraise the chained forecasts with varying spans, to allow for the timing of the forecasts, and to realize any possible gains from the additional information conveyed by quarterly data. The annual benchmark forecasts were computed by averaging the extrapolations for the four quarters of a given year. Much has been learned through work in this area about how extrapolations of different types are related to the predictive span, quality of the data, variability of the series, etc. These lessons are pertinent for the appraisal of forecasts proper, inasmuch as the latter contain elements of extrapolation, and I shall refer to them repeatedly in this chapter.

Many economic time series, and particularly comprehensive aggregates such as GNP, are highly autocorrelated. This implies high correlations among the lagged terms in the N3 equations. The large standard errors of all but a very few of the regression coefficients of these terms reflect the resulting multicollinearity. In quarterly regressions only the two shortest lags appeared significant according to the t-tests of these coefficients. Even tests of the combined significance of several additional terms taken together would not give any definite support for inclusion of earlier terms (longer lags) in the N3 regression.[7] Nevertheless experiments showed that addition of a few longer lags does, on the whole, improve the accuracy of the extrapolations. The R_3 ratios for GNP and industrial production presented in the next section are based on autoregressive equations with five lagged terms (A_{t-i}, $i = 1, 2, \ldots 5$).

This illustrates the basic difficulty of inferring predictive properties

[6] Quarterly figures for GNP and components in the period before World War II are of questionable quality. The use of wartime data raises serious problems that are familiar. Apart from the quality and consistency of the data, there is the basic question of the continuity of the process or processes they measure. The issue here is over what period a sufficiently stable autoregressive structure can be assumed.

[7] These tests used F-ratios of the general form $\frac{1 - \bar{R}^2_{t, t-1, \ldots t-k}}{1 - \bar{R}^2_{t, t-1, \ldots t-n}}$, where $k < n$, to determine the past period A_{t-k} for which the additional set A_{t-k-1} to A_n yields no further increase in $\bar{R}^2$ (the adjusted multiple determination coefficient for the autoregressive equation N3). The results noted in the text were obtained even on standards regarded conventionally as "loose," such as significance levels of .10 and above.

from statistical estimates obtained from historical data. Regression statistics, for example, may often do little more than describe some average relations within the period of fit. Such information is likely to prove insufficient for developing decision rules to guide the forecaster (e.g., rules on how many and which lagged terms to include in a predictive autoregressive model).

This paper relies mainly on N1 and N2*, the last level and the average change extrapolations introduced in Table 1. Model N2 gives inferior results for most of the variables examined. Work with autoregressive models such as N3 yields worthwhile information about the properties of both the time series data and extrapolative predictions, but it still poses serious problems of principle and application.

ANNUAL FORECASTS VS. EXTRAPOLATIONS

Comparisons of forecasts proper with benchmark forecasts derived from the extrapolative models were made by calculating the ratios of the corresponding root mean square errors (M_P/M_N). As noted in the section on bias in Chapter 4, these measures involve squaring the individual forecast errors, which results in greater weights being attached to larger than to smaller deviations from the recorded values.[8] The ratios bear the subscript of the model used in computing the denominator, e.g., $R_1 = M_P/M_{N1}$. Accordingly our four models provide as many ratios: R_1, R_2, R_2*, and R_3.

For annual forecasts of GNP and industrial production, comparisons with all four models are presented. The root mean square errors of the extrapolations for the period 1953–63 are as follows:

	M_{N1}	M_{N2}	M_{N2}^*	M_{N3}
GNP (billion dollars)	24.60	19.34	15.31	15.39
Industrial production (1947–49 = 100)	10.58	11.78	9.91	9.61

Table 16 shows that, for the annual forecasts of GNP and industrial production, the ratios generally are less than unity, indicating that these predictions pass the extrapolative model tests (i.e., $M_P < M_N$).

[8] That is, $M^2_P = \frac{1}{n}\Sigma(P_t - A_t)^2$. More nearly comparable with the other summary measures of predictive accuracy (the simple arithmetic and absolute averages) is the square root of the above figure, M_P. Except in the trivial case where all errors are equal, M_P is always larger than the mean absolute error $|\bar{E}|$ (and the latter is larger than the mean error $\bar{E}$).

TABLE 16

Annual Forecasts of GNP and Industrial Production: Comparisons with Four Extrapolative Models, 1953-63

Line	Forecast Set[a]	Period Covered	Root Mean Square Error,[b] M_P (1)	Root Mean Square Errors: Ratios of Forecast to Extrapolation[c] R_1 (2)	R_2 (3)	R_2^* (4)	R_3 (5)
			Gross National Product				
1.	A	1954-63	12.51	.506	.651	.781	.798
2.	B	1953-63	10.69	.435	.553	.699	.695
3.	C	1958-63	11.04	.424	.534	.797	.686
4.	D	1956-63	11.40	.459	.631	.939	.758
5.	E	1953-63	16.71	.679	.864	1.091	1.086
6.	F	1953-63	8.84	.359	.457	.578	.574
7.	G	1953-63	7.93	.322	.410	.518	.515
8.	H	1954-63	12.05	.487	.626	.752	.768
			Industrial Production				
9.	A	1954-63	5.76	.559	.491	.554	.618
10.	C	1958-63	4.92	.515	.393	.582	.453
11.	D	1954-63	5.28	.512	.450	.508	.566
12.	E	1954-63	5.78	.561	.493	.556	.620
13.	F	1953-63	4.65	.440	.395	.469	.484
14.	G	1953-63	4.83	.457	.410	.487	.503
15.	H	1954-63	4.68	.454	.399	.450	.502

[a]This table covers the same forecast sets as Table 1. It refers to forecasts of levels only. For details, see notes to Table 1.

[b]The general formula for the root mean square error is $M_P = \sqrt{\frac{1}{n} \Sigma \ (P-A)^2}$, where P and A are forecasts and actual values, respectively, and the summation is over the periods covered by the forecasts (see text and footnote 7). Entries in this column are in billion dollars for GNP (lines 1-8) and in index points (1947-49 = 100) for industrial production (lines 9-15).

[c]The ratios are: $R_1 = M_P/M_{N1}$; $R_2 = M_P/M_{N2}$; $R_2{}^* = M_P/M_{N2}{}^*$; and $R_3 = M_P/M_{N3}$. The denominators, M_N, are the root mean square errors of four types of extrapolations. N1 refers to the projection of the last known level, N2 to that of the last known change, N2* to that of the average historical change, and N3 to that of the average relation between the present value of the series and its past values (based on regressions of A_t on A_{t-i}, i = 1, 2, ... 5). See text for more detail.

The measures refer to level forecasts, but the corresponding ratios for change forecasts are similar and the statement applies to them as well.[9]

For GNP, the R_3 ratios are about equal to R_2^* in the periods after 1953–54, and considerably lower in the years after 1956 or 1958. The R_2 ratios are substantially smaller and R_1 still lower. This means that N3 or N2* provide the most stringent standards and N1 the easiest. Judging solely from these measures, forecasters in one group would not have done much worse if they had simply extrapolated the recent trend in the manner of N2*, and in another case such extrapolation would have actually done better (lines 4 and 5).

However, it should be recalled that the changes predicted from year to year by forecasters are generally well correlated with the actual changes (see Charts 1 and 2 and text in Chapter 3). As will be shown in the following section (Chart 3 and text), changes derived by extrapolations show much lower correlations with the recorded changes. This advantage of greater efficiency (higher correlations with ΔA) is not necessarily offset by the disadvantage of greater bias that the forecasts usually have relative to such extrapolations as N2* and N3. While these aspects of forecasters' performance are important, they cannot be revealed by the root mean square errors M, which measure the over-all accuracy of predictions as affected by both bias and efficiency. To deal with the two aspects separately, it is necessary to compare other measures than just the total M figures.

For industrial production, the simplest extrapolations (N1) would have done surprisingly well recently. In fact, the R_1 ratios here are throughout higher than R_2, about equal to R_2^* in the period after 1953–54, and only moderately lower than R_3 (Table 16, lines 9, 11–15). In the latter part of the period covered, R_1 was exceeded only by R_2^* (line 10).

Comparisons in terms of index numbers on a common base showed the mean absolute errors to be, as a rule, somewhat larger for the industrial production forecasts than for the corresponding GNP forecasts (Table 1, lines 9–22, columns 4 and 5). However, the M_P/M_N

[9] The ratios are more meaningful for the levels, since the forecasters' base estimates (ECP) are themselves often derived by, or attributed to, some kind of extrapolation. Where the errors with respect to the base values are smaller for the forecasts than for the extrapolations, the level ratios will typically be more favorable to the forecasts than the change ratios. This is true for the cases mentioned earlier in which the ECP errors make the predictions of levels more inaccurate than those of changes.

ratios in Table 16 tend to be lower for industrial production than for GNP, when compared for the same sources and extrapolative models. This would suggest that, relative to these extrapolations, industrial production was predicted better than GNP. But it is important to note that this result depends on the particular extrapolative procedures adopted. To make the measures for the two variables and for the different models comparable, the models were computed in all cases on the assumption that the last value known to the forecaster is that for the third quarter of the base year. This fits well enough the situation for GNP, a quarterly series, but industrial production is available monthly and those who forecast this variable at the end of the year usually know its preliminary estimate for October and may even know that for November. When models N1, N2, and N2* for industrial production were recomputed, with November as the last known position or base, very considerable improvements of the extrapolations and corresponding increases in the R ratios were obtained, compared with the results shown in Table 16, lines 9–15.[10]

The differences between the extrapolations for GNP and industrial production reflect differences in the behavior of the two series over time. To take the simplest illustration, it is clear that N1 yields smaller errors than N2 at turning points. In general, N1 performs better on the more cyclical and irregular series. N2 performs better only for smooth series with persistent trends. Industrial production fluctuated considerably more than GNP; its growth was somewhat weaker and less steady. This can explain why N2 is better than N1 for the annual predictions of GNP but worse for industrial production.

Again, it is instructive to compare N1 and N2* in this context. The main weakness of the former model is that, in disregarding change, it produces forecasts with a large bias (underestimation errors for the growing series). Model N2* corrects largely for this bias, but it too is worse than N1 on turning points (and also on marked retar-

[10] The R_1 ratios showed the greatest increases, of up to 50 per cent; on the November base, these ratios fall in the .65 to .80 range and considerably exceed the corresponding measures for GNP. The increases in R_2^* ratios were much smaller and those in R_2 still smaller (note that the weight of the base period is smaller in models N2 and N2* than in N1). Among the recomputed ratios, R_1 exceeded the others, except in the most recent years when the R_2^* ratios were higher (the R_3 ratios were not recomputed; our autoregressive extrapolations all use quarterly data only).

dations). As already indicated, the role of the bias is greater for GNP, and that of the turning-point errors for industrial production.[11]

PATTERNS OF ERRORS IN FORECASTS AND EXTRAPOLATIONS

Chart 5 shows for benchmark extrapolations what Charts 1 and 2 showed for the forecasts proper. It is clear that the results for the extrapolations are decidedly inferior.

Consider the model N2*, that is, the projections of average historical change, which turned out to be relatively effective for GNP in terms of the over-all average errors (Table 16). Chart 5 shows that the changes predicted by this model are nearly constant from one year to the next; they approximate a straight line cutting through the "actual" GNP changes at the $20 billion level. This, of course, would be necessarily so for any series with a pronounced and relatively stable growth rate, and the results for industrial production are indeed very similar to those for GNP.

In short, the strength of N2* lies in its being a relatively good trend estimate, thus nearly free of bias. But the weakness of this model lies in its being poorly correlated with the actual changes (in other words, in the high residual variance component of its errors). Especially in periods of large deviations from the trend, as in the recession years 1954 and 1958, the errors produced by N2* must be and obviously are large, as shown in Chart 5, for both GNP and industrial production.

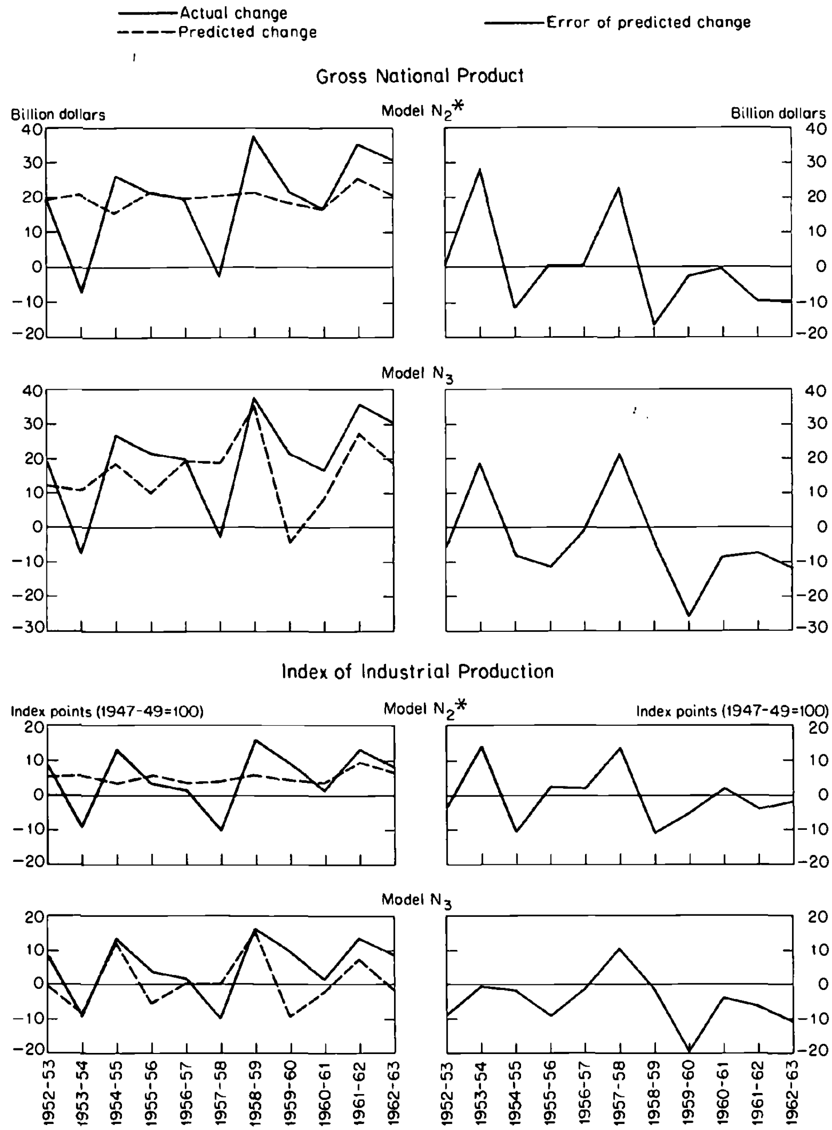
The autoregressive model N3, which uses lags of one to five quarters, yields forecasts that, unlike those of N2*, vary a great deal from year to year. However, the correlation of these forecasts with actual changes is low for GNP and only moderate (though significant) for industrial production. The errors of the two models are similar in size. They also show some correspondence in their year-to-year changes, notably peak values in the 1953–54 and 1957–58 intervals.

Thus it is easy to see that, in terms of correlation with actual

[11] Further evidence along the same lines is provided by results for other variables. Consumption is similar to GNP in smoothness and trend, and for both series the models N2* and N3 rank highest and N1 the lowest. (Five lagged terms are needed in both cases to make N3 about as good as N2*.) For plant and equipment expenditures, a series with more variability and less trend than the others, N3 turns out to be the best of the models within the whole range of one to five lagged terms. Here, as for the production index, N1 ranks higher than N2*, and N2 is the worst.

CHART 5

Two Types of Extrapolations of Industrial Production and GNP, Actual and Predicted Changes and Errors, 1953–63



changes, the annual forecasts of GNP and industrial production are definitely better than either of the extrapolations N2* and N3. The coefficients of correlation between actual and predicted changes in the tabulation below for 1953–63 provide numerical evidence in support of this statement:

	Forecasts			*Extrapolations*	
	E	F	G	N2*	N3
GNP	.642	.900	.882	.182	.424
Industrial production	.753	.920	.895	.535	.601

The fact that the actual and predicted year-to-year changes in GNP and industrial production are in general well correlated should not be interpreted to mean that the changes are effectively forecast a full year ahead. Our analysis of the multiperiod forecasts shows that forecasters typically achieve a considerable measure of success in predicting the next one or two quarters ahead, but limited success beyond that. Let us also recall that the annual record can likewise be interpreted as implying an effective forecasting range of little more than two quarters. Not only are the forecasts typically made late in the preceding calendar year, but patterns of performance very similar to those shown here could be obtained merely by accurately forecasting the first one or two quarters of the ensuing year and assuming no change beyond that. This is indicated by our experiments with *ex-post* extrapolations summarized later in this chapter (see Table 19 and Chart 5).

SECTORAL FORECASTS, EXTRAPOLATIONS, AND BUSINESSMEN'S ANTICIPATIONS

The summary measures of relative error show forecasts of consumption to be better than forecasts of government spending and much better than those of investment (Table 5, column 1). When taken relative to the extrapolative benchmark models, however, consumption often comes out *worse* than either investment or government expenditures (Table 17, columns 2 and 3). In fact, the R_2^* ratios in set F exceed unity only for consumption and residential construction. They are very high for all three of the major consumption components, which indicates the marked inferiority of these forecasts to simple trend extrapolations. Consumption of durables, which is by far the most variable of these components and which has the largest errors of relative change, shows the lowest R_2^* ratio of the three. The corresponding ratios for consumption forecasts B and C are only slightly less than one.

Since consumption has been rising steadily in recent years, the strong showing of the trend extrapolation produced by the N2* model is not difficult to understand. This applies particularly to consumption of nondurables and services. Other GNP components are much less smooth, hence for them such projections are far less effective.

It is clear that the consumption forecasts fail to make use of much information in the past behavior of households that could have significantly improved the forecasts. Consumption may be able to provide the greatest *scope* for improvement of the GNP expenditure forecasts.

Forecasts of construction (especially residential), changes in inventories, and net foreign investment are greatly in need of improvement. Although these series have relatively weak trends and strong cyclical and irregular variations, the N2* extrapolations for them are often nearly as good as the forecasts proper, and sometimes better. This finding will be recognized as particularly unfavorable to the forecasts, since the extrapolations must be rather inefficient here (unlike the case of consumption). The evidence on construction is given in Table 17. The forecasts of net changes in inventories and net foreign investment, as can be seen from the tabulation below, tend to be weak and inferior to the total investment forecasts (the ratios are particularly high for the A forecasts which, it must be remembered, have considerably longer spans than the others). Since these variables assume negative values, the ratios shown are based on absolute change errors.[12]

	R_2^* *Ratios for Forecast Sets*		
	A	B	F
Gross private domestic investment	1.360	.883	.618
Net change in inventories	1.457	1.048	.843
Net foreign investment	1.185	1.017	.992

Forecasts of producers' durable equipment are in most cases better relative to extrapolations than forecasts of consumption, housing, inventory changes, and net foreign investment. The same is true of the forecasts of total plant and equipment outlays, although the record for the plant (or nonresidential construction) component tends to be worse

12 The differences between the ratios for absolute and those for percentage errors tend to be small when both units of measurement are applicable (compare the ratios for GPDI in the above tabulation and in Table 17).

TABLE 17

Forecasts of Relative Changes in Major Components of GNP: Comparisons with Extrapolative Models, 1953-63

Line	Predicted Variable	Root Mean Square Error, M_P (percentage points) (1)	Root Mean Square Errors: Ratio of Forecast to Extrapolation R_1 (2)	R_2* (3)
		Forecast Set B: 1953-63		
1.	Gross national product	2.15	.395	.660
2.	Personal consumption expenditures	1.52	.318	.911
3.	Gross private domestic investment	12.83	.901	1.010
4.	Plant and equipment	7.80	.901	1.127
5.	Housing	14.12	1.229	2.028
6.	Total government expenditures:	2.82	.406	.457
7.	Federal	5.48	.681	.605
8.	State and local	2.01	.239	1.214
		Forecast Set F: 1953-63		
9.	Gross national product	1.70	.312	.521
10.	Personal consumption expenditures	1.69	.354	1.015
11.	Gross private domestic investment	7.20	.506	.567
12.	Total government expenditures	2.15	.309	.348
		Forecast Set F: Other Periods[a]		
13.	Consumer durables	5.36	.663	1.250
14.	Consumer nondurables	1.56	.444	2.425
15.	Consumer services	1.78	.280	1.334
16.	Producers' durables	6.05	.495	.603
17.	Nonresidential construction	4.01	.719	.904
18.	Residential nonfarm construction	9.15	.931	1.596
19.	Federal government expenditures	1.86	.315	.527
20.	State and local expenditures	1.35	.160	.741

(continued)

TABLE 17 *(concluded)*

Line	Predicted Variable	Root Mean Square Error, M_P (percentage points) (1)	Root Mean Square Errors: Ratio of Forecast to Extrapolation R_1 (2)	R_2* (3)
		Forecast Set A: 1958-63		
21.	Gross national product	1.92	.368	1.027
22.	Personal consumption expenditures	1.01	.222	.771
23.	Consumer durables	5.21	.901	.935
24.	Gross private domestic investment	11.43	1.118	1.255
25.	Producers' durables	6.36	.562	.593
26.	New construction	5.08	.898	1.917
27.	Total government expenditures	3.15	.410	.861
		Forecast Set C: 1958-63		
28.	Gross national product	2.04	.374	.726
29.	Personal consumption expenditures	1.44	.310	.960
30.	Gross private domestic investment	8.50	.570	.666
31.	Plant and equipment	4.95	.597	.643
32.	Residential construction	9.04	.865	2.251
33.	Total government expenditures	2.17	.319	.893
34.	Federal	3.56	.593	1.027
35.	State and local	1.94	.232	1.019

[a]Entries on lines 13-15 cover 1959-63; lines 16, 19, and 20, 1955-63; lines 17 and 18, 1956-63.

Note: This table covers the same forecast sets as Table 5 and, like the latter, uses measures based on errors of percentage changes (as defined in Chapter 4, text and note 5). The extrapolations match the forecasts strictly (for set A, they refer to the last quarter of the year wherever the forecasts do, see Table 5, note c; in all other cases, they are annual).

than that for equipment.[13] A marked exception, however, is provided by set B, in which the plant and equipment predictions are decidedly poor, worse than the trend projections and than the predictions for all but two GNP components (see Table 17, column 3).

[13] Compare lines 16 and 17, columns 2 and 3, of Table 17. It should be noted that the opposite is true of the absolute errors; see the corresponding measures in column 1 and also footnote 6 in Chapter 4.

It is possible that these findings reflect the predictive value and usefulness to forecasters of data on investment anticipations. Most of the end-of-year forecasts can and do use the figures from the McGraw-Hill Survey of Investment Intentions, which is conducted annually and released in November. However, company forecast B and also the group forecasts D and E are made in October, that is, before the release of this survey. It is particularly interesting here to compare the performance of sets B and F, which are generally similar in over-all quality. As shown in Table 17, the errors in forecasts of gross private domestic investment were much larger for set B than for set F (the *M* values are 13.4 and 7.5, respectively, see lines 3 and 11, column 1) and most of this difference can be traced to the plant and equipment component of investment.[14] Whereas in forecasting consumption and the rest of GNP other than fixed investment the errors in set B are either smaller than, or about the same as in set F, in predicting GNP as a whole, set F is appreciably better than set B. This is clearly because of its superiority in the investment sector, which very likely reflects the fact that set F did, whereas set B did not, utilize the investment anticipations data.

In this connection, it is also interesting to note that the base-period estimates (ECP) are often considerably better for business capital outlays than for other GNP components. This, too, is probably due in large measure to the availability of anticipations surveys, including the quarterly ones made jointly by the Commerce Department's Office of Business Economics and the Securities and Exchange Commission (OBE-SEC).

The OBE-SEC *annual* surveys are conducted in January–February and published in March; hence, they provide, in effect, shorter-span predictions than those made by our forecasters at earlier dates, between October and January. The predictive record of these survey data is as good as it is partly because of their relatively short horizon. Even so, it is significant that they show much higher correlations with expenditures on producers' durables and nonresidential construction than the global investment forecasts in our collection do.[15] Such com-

[14] Compare the errors of set B in plant and equipment (Table 17, line 4) with the errors of set F in producers' durables and nonresidential construction (lines 16 and 17). The differences in favor of set F persist when the errors are recomputed for 1956–63, the period in which these forecasts overlap.

[15] The forecast and actual values are for the GNP components representing private "fixed investment," that is, the sum of the above two categories of expendi-

parisons, though inevitably crude, provide some additional evidence of the relative success of anticipations data for fixed investment.[16]

FORECASTS AND EXTRAPOLATIONS OVER VARYING SPANS

While average forecast errors nearly always increase with the lengthening of the predictive span, the R_1 ratios show no definite tendency to rise along with the span (Table 18). This is so because the errors of base-level extrapolations also grow larger as the period over which the extrapolations are made is lengthened. The same applies to the R_2 ratios, which are not shown in the table (they are, in most cases here, smaller than R_1, contrary to the case for the annual GNP forecasts).[17]

On the whole, the R_1 ratios do not vary greatly for the different spans. In some cases, particularly for industrial production, they do increase appreciably when applied to longer forecasts (Table 18, lines 23 or 35). Elsewhere, counterexamples are found, however: a longer forecast, despite its larger average error, may be better than a shorter one when judged by comparison with a naive model (lines 8, 11, and 17). The ratios are all less than one, that is, the errors of extrapolations N1 increase with the span sufficiently to remain on the average larger than the errors of the forecasts proper.

The R_2^* and R_3 ratios are all considerably larger than either R_1 or R_2. When two lagged terms are used in the N3 autoregressions, the results tend to be inferior to those of the trend projections N2* (i.e., $R_3 < R_2^*$ in most cases). But the differences between these ratios are on the whole small, and inclusion of additional lags should here and

ture. The anticipations data are based on a somewhat different concept which excludes private nonbusiness institutions, capital outlays charged to current expense, and some other items. In terms of levels, therefore, forecasts have smaller errors than anticipations when both are compared with the investment components of GNP; but this, of course, merely reflects differences in measurement, not in predictive performance. What is meaningful, however, is that the anticipations data (Z) show higher correlations with the recorded fixed investment outlays (I) than the global forecasts (P) do. Moreover, forecasts hardly add anything to a statistical explanation of the variance of I after allowing for the high correlation between I and Z. The partial correlation coefficients $r_{IP.Z}$ are in all cases small absolutely and a few are negative, while the values of $r_{IZ.P}$ are all positive and relatively high.

[16] Several recent writings offer substantial evidence that the expectations or intentions of business management regarding fixed investment have direct and considerable predictive value. See, e.g., the essays in *The Quality and Economic Significance of Anticipations Data* (Princeton for the National Bureau of Economic Research, 1960), particularly those by Dexter M. Keezer, *et al.* (pp. 369–386), and Arthur M. Okun (pp. 407–460).

[17] When shorter time units are used, as in the chained forecasts, the advantage that N2 has for the smoother annual series is largely lost.

TABLE 18

Forecasts of GNP and Industrial Production: Comparisons with Two Extrapolative Models over Spans from Three to Eighteen Months, 1947-63

Line	Forecast Set[a]	Period Covered	Error Statistics[b]	Span of Forecast (months) Three (1)	Six (2)	Nine (3)	Twelve (4)	Fifteen (5)	Eighteen (6)
				Gross National Product					
1.	A	1947-49, 1955-56, 1958-63	M		15.3		24.2		
2.	A	1947-49, 1955-56, 1958-63	R_1		.680		.752		
3.	A	1947-49, 1955-56, 1958-63	R_{2*}		1.107		1.568		
4.	C[c]	1958-63	M	7.5	10.8	12.3	15.0	19.5	
5.	C[c]	1958-63	R_1	.412	.456	.413	.417	.428	
6.	C[c]	1958-63	R_{2*}	.610	.793	.904	1.086	1.127	
7.	D	1956-63	M		12.6		18.1		
8.	D	1956-63	R_1		.549		.527		
9.	D	1956-63	R_{2*}		.983		1.136		

(continued)

TABLE 18 *(continued)*

Line	Forecast Set[a]	Period Covered	Error Statistics[b]	Span of Forecast (months) Three (1)	Six (2)	Nine (3)	Twelve (4)	Fifteen (5)	Eighteen (6)
				Gross National Product					
10.	D	1959-63	M	8.2	12.8	13.9	15.9		
11.	D	1959-63	R_1	.397	.445	.421	.392		
12.	D	1959-63	R_{2*}	.763	.874	.962	1.043		
13.	E	1956, 1960-63	M		11.9		23.0		
14.	E	1956, 1960-63	R_1		.475		.582		
15.	E	1956, 1960-63	R_{2*}		1.310		1.807		
16.	G[d]	1955-63	M	8.4	10.0	13.8	15.4	17.1	17.0
17.	G[d]	1955-63	R_1	.520	.445	.552	.534	.571	.504
18.	G[d]	1955-63	R_{2*}	.764	.680	.874	.920	1.040	1.016

(continued)

TABLE 18 *(continued)*

Line	Forecast Set[a]	Period Covered	Error Statistics[b]	Span of Forecast (months) Three (1)	Six (2)	Nine (3)	Twelve (4)	Fifteen (5)	Eighteen (6)
				Industrial Production					
19.	A	1947–49, 1955–56, 1958–63	M		6.4		7.0		
20.	A	1947–49, 1955–56, 1958–63	R_1		.663		.651		
21.	A	1947–49, 1955–56, 1958–63	R_{2*}		.792		.780		
22.	C[c]	1958–63	M	4.4	5.6	6.4	7.0	9.7	
23.	C[c]	1958–63	R_1	.529	.571	.592	.605	.681	
24.	C[c]	1958–63	R_{2*}	.546	.662	.833	.997	1.266	
25.	D	1947–63	M		5.0		9.0		
26.	D	1947–63	R_1		.708		.801		
27.	D	1947–63	R_{2*}		.670		.815		

(continued)

TABLE 18 *(concluded)*

Line	Forecast Set[a]	Period Covered	Error Statistics[b]	Span of Forecast (months) Three (1)	Six (2)	Nine (3)	Twelve (4)	Fifteen (5)	Eighteen (6)
				Industrial Production					
28.	D	1959–63	M	4.8	6.4	8.2	7.1		
29.	D	1959–63	R_1	.577	.600	.640	.527		
30.	D	1959–63	R_{2*}	.613	.710	.794	.752		
31.	E	1951–63	M		8.0		7.8		
32.	E	1951–63	R_1		.889		.749		
33.	E	1951–63	R_{2*}		.790		.655		
34.	G	1956–63	M	4.0	4.9	8.1	9.7	10.4	9.6
35.	G	1956–63	R_1	.678	.605	.818	.858	.839	.678
36.	G	1956–63	R_{2*}	.684	.620	.853	.924	.954	.800

Notes to Table 18

[a]All measures refer to level forecasts. The forecast sets are the same as those in Table 9, with two exceptions: (1) for GNP, a subset of D was added covering the years 1953-63, with sixteen observations per span (lines 7-9); (2) for industrial production, another subset of D was added covering the years 1947-63, with thirty-three observations per span (lines 25-27). For further details, see Table 9.

[b]The meaning of the Symbols is as follows:

$M = \sqrt{\frac{1}{n} \Sigma (P-A)^2}$, which is the root mean square error of forecast (in billion dollars for GNP; in index points, 1947-49 = 100, for industrial production).

$R_1 = M/M_{N1}$; $R_2 = M/M_{N2*}$. These are the root mean square error ratios of forecast to extrapolation.

[c]The entries on this line are not strictly comparable because some of the forecast chains are "incomplete," so that the number of observations for different spans varies. See Table 9, note b.

[d]Based on forecasts in constant prices, as reported, and on corresponding extrapolations.

there tip the scales in favor of N3 (the evidence on this point is still incomplete).

In most cases the R_2* ratios show increases when the predictive span is extended (Table 18). This means that the N2* extrapolations do not deteriorate as much as the forecasts proper, so that their relative performance improves as spans lengthen. It should be noted that the same applies to the autoregressive predictions: the R_3 ratios also tend to rise for the longer forecasts. Model N3 does indeed suggest the likelihood of this result.[18]

[18] Suppose that $A_t = a + b\,A_{t-1} + u_t$. Then $A_{t+1} = a + b\,A_t + u_{t+1} = a(1+b) + b^2A_{t-1} + (bu_t + u_{t+1})$. Over the span of one period, the variance of the residuals is var (u_t); over two periods, it is var $(bu_t + u_{t+1})$ which, assuming that the u's are independent and have equal variances, equals $(1 + b^2)$ var (u). Thus the variance increases with the span, depending on the size of b. This is illustrated here in the simplest case of a first-order autoregression, but the same argument applies to more complex situations.

Extrapolations for increasing spans are derived by stepwise "chaining" of predicted and observed values as follows. Let the forecast for one quarter ahead be $\hat{A}_t = a + b_1A_{t-1} + b_2A_{t-2} \cdots$; for two quarters, $\hat{A}_{t+1} = a + b_1\hat{A}_t + b_2A_{t-1} \cdots$; etc. The coefficients a, b_1, $b_2 \cdots$ are estimated in the first step (which may

For two sets of GNP forecasts, the R_2^* ratios are all larger than one (lines 3 and 15), indicating that here both the six- and the twelve-month predictions are on the whole worse than the average change extrapolations over the periods covered.[19] For other sets, the R_2^* ratios are less than one for the short spans, but they approach or exceed one for the longer spans (mostly nine to fifteen months, in one case fifteen to eighteen months; see Table 18, lines 6, 12, and 18).

For industrial production, the ratios grow smaller as the spans lengthen. Our table includes only three instances of $R_2^* > 0.9$ and only one of $R_2^* > 1.0$, all in the range of twelve- to fifteen-month forecasts (lines 24 and 36).

Presumably, the observed association between longer forecasts and larger errors can be partly explained by the fact that extrapolations of various types also tend to worsen as the predictive span lengthens. The past behavior of the series to be predicted will usually be considered in some way by the forecaster, and many explicitly use various extrapolative techniques. Forecast and extrapolation often have much in common, and the latter may be viewed as one of the ingredients of the former.[20] However, forecasters apparently fail to use the historical content of the series as efficiently as they might even by fairly simple and inexpensive means, perhaps because they pay too much attention to the very few values in the most recent past. Better use of trend projections would have improved many forecasts, particularly those of GNP over longer spans. This can be inferred from the comparisons of forecasts with the N2* model. As shown earlier, the growth of the economy has often been underestimated in the forecasts, an error that simple trend extrapolations would have helped to reduce.

---

relate to the fourth quarter of the base year, ECP) and retained in the following steps for the given multiperiod forecast (say, for the estimation of the four quarters of the next year).

Some experiments were also made with another approach, consisting of separate fittings for different spans: regressions of $A_t, A_{t+1}, \ldots A_{t+4}$ on $A_{t-1}, A_{t-2} \ldots$ (this gives five equations for each year covered by the predictions). It is interesting that the two approaches gave very similar results for GNP. In general, however, the chained forecasts are more accurate than the fits by span.

[19] As will be shown later, the results of such comparisons depend on the periods covered. It should be noted that one of these sets (A) includes the early postwar years, 1947–49, for which forecasts were particularly poor.

[20] Statistical methods, e.g., partial correlation analysis, can help evaluate this relation; work along these lines is in process (see footnote 2 above).

Other ingredients of forecasts may, to be sure, also contribute to the decreasing accuracy of longer-range predictions. These ingredients —variables that typically move in advance of the predicted series, anticipations data, and the forecasters' own judgment—also are probably decreasingly reliable over longer spans.

NEAR-TERM FORESIGHT AND ACCURACY OF ANNUAL PREDICTIONS

We find, then, that forecasts of the near future are definitely superior to all types of extrapolation, while forecasts with longer spans, of nine or twelve months and more, are not (they are in fact worse than some extrapolations). Let us also recall that the year-to-year forecasts proved generally better than extrapolations. Are these findings consistent? Does the record of the annual forecasts imply greater accuracy than that of the forecasts with varying spans?

The answers are yes to the first question and no to the second. The annual forecasts are generally made late in the preceding calendar year. They can be viewed as averages of the quarterly or semiannual forecasts and some are actually so computed; their mean spans, then, are roughly six months (or again somewhat more than that for the "effective span"). Errors of the (explicit or implicit) forecasts for parts of the year could in some degree be mutually offsetting, which would tend to make the annual forecasts better than most of the shorter ones. This seems, indeed, to be true in some cases. On the whole, however, annual forecasts definitely tend to be better only relative to the forecasts for the late parts of the year, not to those for the early parts.

Furthermore, a good predictive record for the first two quarters would be sufficient to produce a moderately good record for the year as a whole. Even the knowledge of the first quarter alone would make a substantial contribution to the quality of annual forecasts of GNP. This is indicated by some experiments we have performed following a suggestion by Geoffrey Moore. These show, in effect, that if one knew what the early parts of the next year would bring and extrapolated this information in simple ways, the resulting annual forecasts would compare very well indeed with the actual forecasts we have examined.

We proceed by constructing hypothetical "*ex-post* forecasts" incorporating information about parts of the predicted future, which is regarded as free of errors. The first assumption is that the forecaster has full knowledge of the level of the series in the fourth quarter of the base year (ECP) and in the first quarter of the coming year. Three types of extrapolations are then applied successively. (1) The level of the series next year is taken simply as equal to the first quarter's level expressed at the annual rate. This extrapolation, a variant of N1, will be called XQ_1. (2) It is assumed that the level in the second quarter will differ from that in the first by the amount of the last "known" change (which here is that from the fourth quarter of the current year to the first quarter of the next year); also, that the third quarter will differ from the second, and the fourth quarter from the third, by the same amounts. This leads to a variant of the N2 model, to be denoted as $XQ_{\Delta 1}$. (3) Incorporating the assumed knowledge of the current and coming quarter, the mean historical change in the series is computed and projected to obtain the predictions of the three remaining quarters of the next year. This is a variant of the N2* model and it is labeled $XQ_{\Delta 1}*$.

The second assumption about the available degree of foresight is that, in addition to ECP and the first quarter, the level of the series in the second quarter of the coming year is also known. Again, the three types of extrapolations are applied, in analogy to N1, N2, and N2*, respectively, which yields the following models. (4) The series is assumed to remain at its second quarter's level in both the third and the fourth quarter (XQ_2). (5) The series is assumed to change in the third quarter by the ("known") amount of its change in the second quarter, and the change in the fourth quarter is also set as equal to the same amount ($XQ_{\Delta 2}$). (6) The average change in the series is computed from the past record, including in addition the changes in the first half of the coming year, the knowledge of which is imputed to the forecaster. The projection of this amount of change over the second half of the year is used to construct annual predictions $XQ_{\Delta 2}*$.[21]

[21] Let the values of a series in quarter IV of the current year and in quarters I and II of the predicted year, expressed at annual rates, be A_0, A_1, and A_2. Let the average historical change incorporating quarters IV and I be $\overline{\Delta A}(1)$ and the corresponding measure incorporating quarters IV, I, and II be $\overline{\Delta A}(2)$. Then the six experiments can be described as follows:

(1) $$XQ_1 = A_1$$

(2) $$XQ_{\Delta 1} = \frac{4A_1 + 6(A_1 - A_0)}{4} = \frac{5A_1 - 3A_0}{2}$$

The results of applying these models of *ex-post* projections to both GNP and industrial production are summarized in Table 19. As shown there, three forecasts covering the years 1953–63 managed to do better than the simple last-level and last-change extrapolations of the coming year's first quarter, XQ_1 and $XQ_{\Delta 1}$ (compare lines 1 and 2 with lines 8–10). For industrial production, the forecasts are even somewhat better than the trend extrapolation $XQ_{\Delta 1}{}^*$, but for GNP the reverse is true (lines 3 and 8–10).

Those extrapolations that use the second quarter as well are generally superior to the forecasts of GNP. For industrial production, XQ_2 and $XQ_{\Delta 2}{}^*$ also score better than the forecasts, but the differences are smaller here and at least one forecast comes out ahead of $XQ_{\Delta 2}$ (compare lines 4–6 and 8–10).

The trend extrapolations tend to be better than the other models, as would be expected (one exception is that XQ_2 gives the best results for industrial production). The only case of decidedly poor performance is $XQ_{\Delta 1}$ for GNP (see lines 1–6).

As a next step, it seemed desirable to relax the strong assumption of perfect knowledge which underlies these models. We replace the actual values for the early part of the predicted year by reported quarterly forecasts and again use extrapolations for the rest of the year. Table 19 shows that such a combination ($GXQ_{\Delta 2}{}^*$), which uses forecasts for the first two quarters and trend projections for quarters III and IV, yields on the average somewhat smaller errors than the forecasts proper for GNP (compare lines 7 and 8–10). In the case of industrial production, the results are mixed: one forecast is a little worse and one better than $GXQ_{\Delta 2}{}^*$.

Chart 6 presents the year-by-year record of predicted changes and their

---

(3) $$XQ_{\Delta 1}{}^* = \frac{4A_1 + 6\,\overline{\Delta A}(1)}{4} = A_1 + \frac{3}{2}\,\overline{\Delta A}(1)$$

(4) $$XQ_2 = \frac{A_1 + 3A_2}{4}$$

(5) $$XQ_{\Delta 2} = \frac{A_1 + 3A_2 + 3(A_2 - A_1)}{4} = \frac{3A_2 - A_1}{2}$$

(6) $$XQ_{\Delta 2}{}^* = \frac{A_1 + 3A_2 + 3\,\overline{\Delta A}(2)}{4}$$

TABLE 19

Comparisons of Annual Ex-Post *Extrapolations and* Ex-Ante *Forecasts of GNP and Industrial Production, 1953-63*

Line	Type of Extrapolation or Forecast[a]	Gross National Product (billion dollars)			Industrial Production (1947-49 = 100)		
		Mean Absolute Error, $\overline{\lvert E \rvert}$ (1)	Mean Arithmetic Error, $\overline{E}$ (2)	Root Mean Square Error, M (3)	Mean Absolute Error, $\overline{\lvert E \rvert}$ (4)	Mean Arithmetic Error, $\overline{E}$ (5)	Root Mean Square Error, M (6)
	Full knowledge of I Q_{t+1}						
1.	XQ_1	8.6	-8.6	10.1	4.6	1.7	5.3
2.	$XQ_{\Delta 1}$	21.3	8.3	22.8	11.1	-0.3	4.9
3.	$XQ_{\Delta 1}{}^*$	4.7	0.1	5.6	4.5	1.5	5.0
	Full knowledge of I & II Q_{t+1}						
4.	XQ_2	4.1	-3.2	5.0	2.3	-0.1	2.8
5.	$XQ_{\Delta 2}$	3.3	1.4	4.4	3.5	0.8	4.2
6.	$XQ_{\Delta 2}{}^*$	3.5	0.7	4.0	2.5	0.9	3.2
	Combined forecast-projection						
7.	$GXQ_{\Delta 2}{}^*$	6.0	3.2	7.0	3.5	1.9	4.1

(continued)

TABLE 19 *(concluded)*

Line	Type of Extrapolation or Forecast[a]	Gross National Product (billion dollars)			Industrial Production (1947-49 = 100)		
		Mean Absolute Error, $\overline{\lvert E\rvert}$ (1)	Mean Arithmetic Error, $\overline{E}$ (2)	Root Mean Square Error, M (3)	Mean Absolute Error, $\overline{\lvert E\rvert}$ (4)	Mean Arithmetic Error, $\overline{E}$ (5)	Root Mean Square Error, M (6)
	End-of-year forecasts[b]						
8.	B	8.4	-1.3	9.7	n.a.	n.a.	n.a.
9.	F	6.3	-2.3	7.5	2.8	-0.7	3.6
10.	G[c]	7.1	3.0	7.9	3.6	1.9	4.3

[a]The models used are as follows (see text and note 21 for full explanation):

Lines 1-3: Assumes knowledge of the values of GNP or industrial production in the first quarter of the predicted year.

Lines 4-6: Assumes knowledge of the values of GNP or industrial production in the first and second quarters of the predicted year.

Line 7: Forecasts G for the first and second quarters combined with average-change extrapolations N2* for the third and fourth quarters.

[b]Measures apply to forecasts of changes, which are appropriate for these comparisons. The extrapolations assume perfect knowledge of the base-period values, hence contain no ECP errors.

[c]Made typically in terms of base-period prices; here converted to current dollars (see Table 1, note d).

CHART 6

Four Types of Ex-Post *Extrapolations of GNP and Industrial Production, Actual and Predicted Changes and Errors, 1953–63*

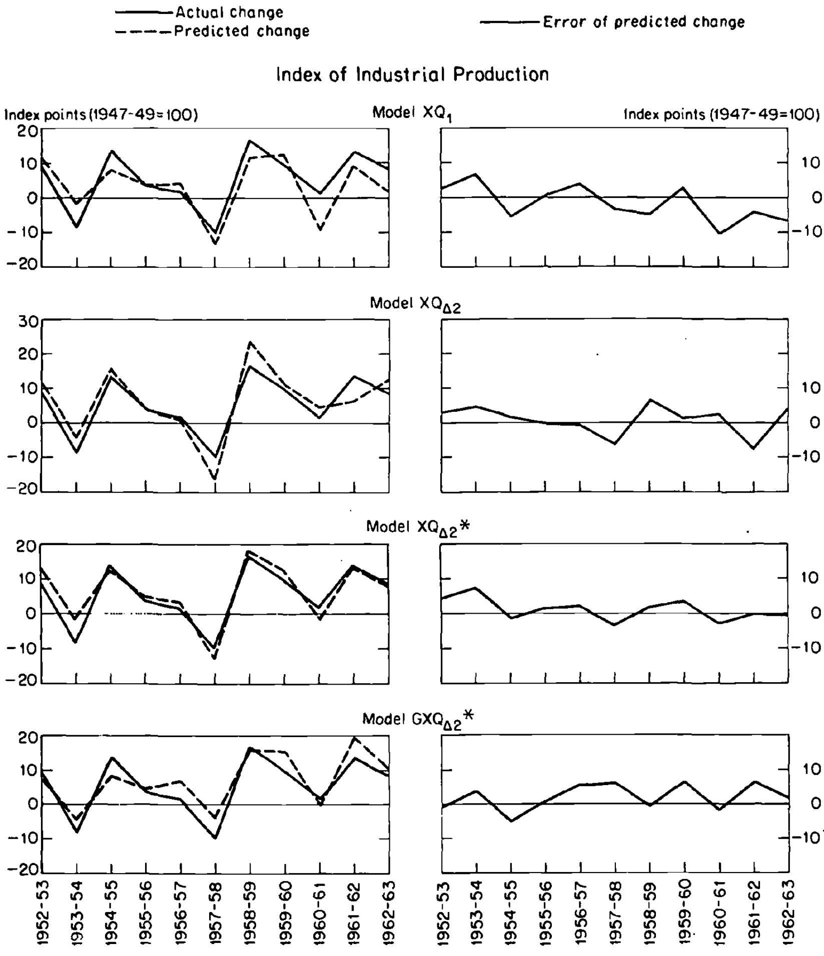
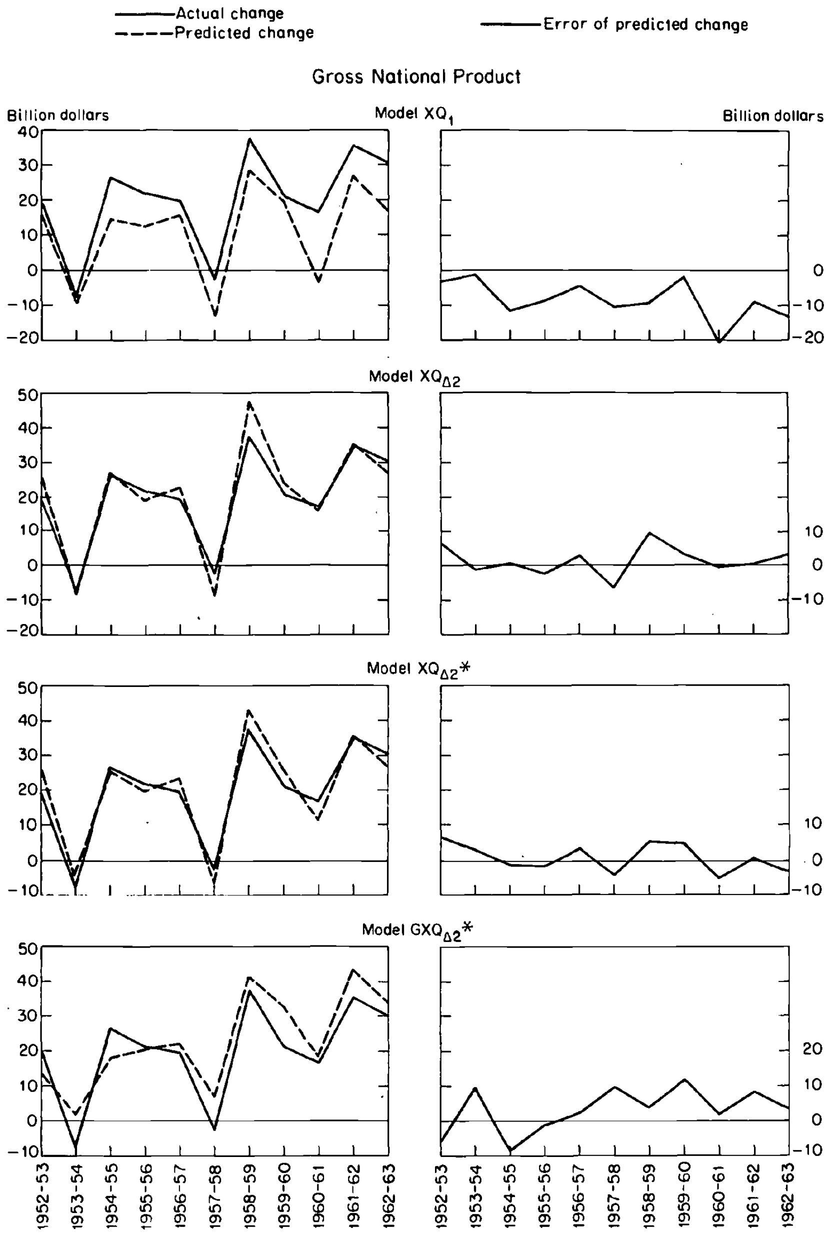


CHART 6 (*concluded*)



errors for four of our seven experimental models.[22] This chart should be compared with Charts 1 and 2 (for the forecasts proper) and Chart 5 (for the *ex-ante* extrapolations).[23] In this way, one can observe that some similarities exist between forecasts and the first-quarter extrapolations, and that great improvements are effected by the incorporation of near-term foresight in the extrapolative models. The most dramatic improvements, in terms of bias elimination as well as reduction of the size of errors, come with the inclusion of the second quarter (which receives a dominant weight in the corresponding models, see the formulas (4)–(6) in footnote 21). The error-increasing and partly biasing effects of using forecasts instead of actual values for quarters I and II can be seen by comparing graphs for $XQ_{\Delta 2}^*$ and $GXQ_{\Delta 2}^*$ in Chart 6. The mild improvements resulting from the use of simple trend projections N2* instead of forecasts for quarters III and IV can only be detected with difficulty by comparing the graphs for $GXQ_{\Delta 2}^*$ in Chart 6 with those for the forecast set G in Charts 1 and 2.

SECTORAL FORECASTS AND EXTRAPOLATIONS FOR DIFFERENT SPANS

Table 20 compares forecasts of GNP components in set C with simple level and trend extrapolations. The average absolute errors of these forecasts were presented earlier (see Table 15).

The ranks of the different variables according to the over-all errors of percentage changes (M_P) are relatively independent of the span. Among the shortest as well as the longer forecasts, total consumption has the smallest errors, government expenditures and consumer durables follow, and the investment variables show the largest errors.[24]

Comparisons with extrapolations yield less consistent results. According to the R_1 ratios, which are based on simple level projections, total

[22] For the corresponding summary measures, see Table 19, lines 1 and 5–7. The models not shown in the chart (they are listed on lines 2–4) include extreme cases such as $XQ_{\Delta 1}$, which works relatively well for the production index but poorly for GNP, and XQ_2, for which the reverse is true. The selection was found sufficiently illustrative for our present purpose.

[23] To facilitate such comparisons, all graphs for GNP were drawn on the same scale, and so were all graphs for the production index.

[24] The values of M_p are larger for residential construction than for plant and equipment, but they are larger still for total GPDI, which suggests a major weakness of the inventory change forecasts. (The latter and the net foreign investment forecasts cannot be included in these comparisons, which refer to relative change errors.)

TABLE 20

Forecasts of Eight Selected Components of GNP: Comparisons with Extrapolations over Spans from Three to Eighteen Months, 1958-63

		Span of Forecast (months)				
Line	Error Statistics[a]	Three (1)	Six (2)	Nine (3)	Twelve (4)	Fifteen (5)
		Personal Consumption Expenditures[b]				
1.	M_P (percentage points)	0.81	1.12	1.40	1.70	2.15
2.	Ratio R_1	0.579	0.444	0.371	0.348	0.321
3.	Ratio R_2*	1.052	1.047	.881	1.156	.964
		Consumer Durables[c]				
4.	M_P (percentage points)	4.55	5.01	5.56	5.77	9.20
5.	Ratio R_1	0.887	0.977	0.788	.996	0.830
6.	Ratio R_2*	0.950	1.029	0.952	1.607	1.309
		Gross Private Domestic Investment[b]				
7.	M_P (percentage points)	6.22	11.00	12.14	14.48	18.51
8.	Ratio R_1	0.772	0.853	0.796	1.059	0.836
9.	Ratio R_2*	0.784	0.890	0.875	1.355	1.014
		Plant and Equipment Outlays[b]				
10.	M_P (percentage points)	3.10	4.77	6.62	8.41	10.52
11.	Ratio R_1	0.856	0.810	0.916	1.031	1.224
12.	Ratio R *	0.852	0.840	1.027	.825	2.234
		Residential Construction[b]				
13.	M_P (percentage points)	5.13	8.91	8.62	7.97	8.46
14.	Ratio R_1	0.902	0.953	0.721	0.595	0.494
15.	Ratio R_2*	0.914	1.009	0.845	0.755	0.636
		Government Expenditures[b]				
16.	M_P (percentage points)	1.43	1.72	2.05	3.00	2.55
17.	Ratio R_1	0.641	0.446	0.367	0.398	0.281
18.	Ratio R_2*	0.734	0.892	0.894	0.930	1.002

(continued)

TABLE 20 *(concluded)*

Line	Error Statistics[a]	Span of Forecast (months) Three (1)	Six (2)	Nine (3)	Twelve (4)	Fifteen (5)
		Net Change in Inventories[d]				
19.	M_P (billion dollars)	2.54	5.31	5.67	6.78	6.58
20.	Ratio R_1	0.692	0.881	0.901	1.292	0.903
21.	Ratio R_2*	0.690	0.873	0.889	1.317	0.899
		Net Foreign Balance[b]				
22.	M_P (billion dollars)	1.09	1.35	1.38	1.51	1.73
23.	Ratio R_1	1.072	0.928	0.766	0.686	0.736
24.	Ratio R_2*	1.048	0.894	0.730	0.659	0.721

[a]M_P denotes the root mean square error of forecast; R_1 and R_2 denote the root mean square error ratios of forecast to extrapolation (see the formulae in Table 18, note 6).

[b]Forecasts cover the period 1958-63. The numbers of observations are 21, 20, 19, 13, and 7 for spans of 3, 6, 9, 12, and 15 months, respectively. The numbers vary because some of the forecast chains do not include all spans (see Table 9, note b).

[c]Forecasts cover the period 1961-63. The numbers of observations for spans of 3 to 15 months are 13, 11, 10, 7, and 4.

[d]Forecasts cover the period 1958-63. The numbers of observations for spans of 3 to 15 months are 19, 17, 16, 10, and 5.

consumption again has the best record (lowest ratios), followed by governmental expenditures; but consumption of durable goods performs no better here than the investment variables. In agreement with the findings for the annual figures, comparisons with simple trend extrapolations are very unfavorable to the consumption forecasts. Most of the R_2* ratios exceed one for both total consumer expenditures and spending on consumer durables, and this includes the short as well as the longer forecasts (Table 20, lines 3 and 6). In fact, these ratios tend to be larger for consumption than for any other major GNP component.

The R_2^* ratios are generally high; 35 per cent of them exceed unity (there is at least one case of $R_2^* > 1.0$ for each of the variables); 25 per cent fall into the range 0.80–0.99; and the remaining 20 per cent are all above 0.6.

The R_1 ratios are as a rule smaller, but they are often sizable. While no more than one-eighth of them exceed 1.0, only about one-fourth are less than 0.6.

The margins by which R_2^* exceed R_1 are particularly large for series with strong trends such as consumption, as would be expected. By the same token, the two ratios are closely similar for series in which trends are unimportant, such as net change in inventories and net foreign balance. Also, the margins in favor of R_2^* often increase with the span of forecast—which presumably reflects the fact that trends are more important over longer periods of time.

Table 20 shows a tendency for the R_2^* ratios to increase with the predictive span in most cases, but not always. Where the ratios rise, they do so irregularly (the only exception is the forecast span of government spending on line 18). For two variables, residential construction and net foreign balance, the ratios are actually higher for the short forecasts than for the longer ones (lines 15 and 24).

There is no reason to expect uniformity in the relation between forecasts and extrapolations for different spans when comparing results for variables with very different characteristics. Such comparisons can be instructive, but additional data are needed at this point and they are hard to get. The available evidence leaves much to be desired, since forecasts C cover a short period of time and their number varies for different spans (see notes to Table 20).

TURNING-POINT ERRORS AND EXTRAPOLATIONS

The use of extrapolative criteria in the appraisal of forecasts has certain implications for turning-point errors. To reconsider the latter, the concepts and notation introduced in the last section of Chapter 4 must be recalled.

A forecast that never predicted any turning points would, of course, have a zero score on "hits" ($TT = 0$), but also a zero score on false signals ($NT = 0$).[25] On the other hand, by missing all the actual turns, such a forecast would have the worst possible score on errors of the second kind: its $\bar{E}_{T2}$ ratio $(= \frac{TN}{TN + TT})$ would equal one. Benchmark forecasts based on extrapolations of the most recent levels or changes or of trends (N1, N2, N2*) do not predict turning points and have, therefore, exactly these characteristics.

The other extreme is represented by a model that predicts a turning point on each occasion. Such a forecast would miss no turns ($TN = 0$) but would have the largest possible number of false signals. Here NT would equal the number of all periods in which there were no observed turns, and the $\bar{E}_{T1}$ ratio would be correspondingly at the maximum.

The built-in disadvantage of forecasts vis-à-vis extrapolations with respect to false warnings may or may not be outweighed by the potential advantage of making correct turning-point predictions. It is difficult to decide what weights to attach to the two kinds of error without information on costs of missed turns vs. false signals to the forecast user.

One simple criterion, implying equal weights, is that the total number of errors of either kind ($n_e = NT + TN$) be less than the total number of turns recorded ($n_a = TT + TN$). The requirement that $n_e < n_a$ reduces to $NT < TT$. In simple extrapolations, of course, $n_e = n_a$, because $NT = 0$ and $TN = n_a$.[26]

However, it could be argued that forecasts for which $n_e = n_a$ may be better than extrapolations, even if they show no fewer errors of both kinds, if they correctly predict at least some turning points (since then $TT > 0$, whereas for the extrapolations TT is always zero).

For ten of the forecast sets listed in Table 8, the condition $n_e < n_a$ is satisfied. In the five remaining cases, $n_e = n_a$.

All measures based on the proportion of errors in the turning-point forecasts, that is, on some combinations of $\bar{E}_{T1}$ and $\bar{E}_{T2}$, are of course functions of TN, NT, and TT only. They disregard NN, the fre-

[25] Thus all terms in the ratio of false signals, $\bar{E}_{T1} = \frac{NT}{NT + TT}$, would in this case equal zero, making the expression indeterminate.

[26] In the model which predicts a turn on each occasion, $n_e = NT$ and $n_a = TT$ (since $TN = NN = 0$). The criterion $n_e < n_a$, therefore, simply describes for this model the requirement that successes be more numerous than failures. The model will meet this requirement if and only if $n_a > 1/2n$, where n is the total number of forecasts $(= n_a + n_e)$.

quency of those periods in which neither the predicted nor the actual values showed a change in direction. The simplest measure that would take NN into account is $\bar{E}_T = \frac{NT + TN}{n} = \frac{n_e}{n_e + n_s}$, where n_s is the number of all directionally correct predictions $(= NN + TT)$ and n the number of all predictions covered. The $\bar{E}_T$ ratios for the forecasts in Table 8 range from 0 to 0.24 and average about 0.12 (that is, the over-all proportion of turning-point error is 12 per cent).

The proportion of turning points observed in the past, n_a/n, is known to the forecaster but is of little help to him, even if it is accepted as the best forecast of the proportion that will prevail in the future. This expected value is a fraction, but the forecaster must decide on each occasion whether or not a directional change is about to occur: he can either predict one or not but he cannot predict a fraction. Suppose that he adopts the following decision rule: never predict a turn for a series which shows $n_a/n < 1/2$ and always predict a turn for a series which shows $n_a/n > 1/2$. This would amount to the use of the first of our two extreme benchmark models in the former case and of the second one in the latter case. But the proportions of predicted turns would then be 0 and 1, respectively, and they would probably be poor approximations to the true fraction n_a/n in either case.

Clearly, it is not satisfactory to compare the forecasts with these limiting alternatives only. An extrapolative benchmark model is needed, which could produce turning points depending on the configurations of previous increases and decreases in the series concerned. Autoregressive models which incorporate several lagged terms can in principle meet this need.

The results of the annual N3 extrapolations for GNP and industrial production are, however, distinctly unfavorable with respect to predictions of the direction of change, as shown for the five-lag models in Chart 3. For GNP, no declines were indicated by the model in either 1954 or 1958, but a slight decline was predicted for 1960, a year in which GNP flattened off in another mild recession but did not fall relative to the preceding year. This record containing two missed turns and one false signal (in comparison with the early estimates) is much worse than that of the forecasts proper, which show no more than one missed turn in the same period. And for industrial production the performance of the model shown in Chart 3 was still weaker. Here the model predicted negative changes in six of the eleven years but only once correctly (for 1954). Again, failures of this magnitude are in general not observable among the forecasts proper.

Evidence on these and other comparisons between forecasts and extrapolations is presented in Table 21. Here the over-all percentages of directional errors ($\bar{E}_T$) are listed for several sets of reported forecasts and constructed benchmark predictions relating to some of the major expenditure series as well as to total GNP and industrial production. It turns out that the performance of N3 with several lagged terms is in most cases inferior to that of other extrapolative models, such as N2* and N3 with but one or two lags (the reverse is true only for plant and equipment outlays; compare columns 1–4, lines 11–15). The addition of lagged terms in the N3 model appears to often have the effect of causing "extra" turns, which both N2* and N3, with fewer lags, avoid.

Comparing the results obtained for different variables, we find that they suggest the following. (1) The frequency of turning-point errors tends to be considerably larger in extrapolations than in forecasts for GNP and, particularly, for industrial production. (2) Directional errors are much more numerous in predictions of gross private domestic investment, and this applies to both forecasts proper and extrapolations. There is some evidence here that forecasters, by and large, did a little better than the adopted benchmark models, but it is certainly not conclusive. For plant and equipment outlays, for example, some of the evidence points in the opposite direction (see lines 4, 5, 13, and 14). (3) Forecasts of consumption show very few turning-point errors, and most sets show no such errors at all. The trend and autoregressive extrapolations N2* and N3 produced no errors of direction for this series which, it will be recalled, was fairly smooth and growing from year to year in the period concerned (see lines 2 and 12).

Some other measures pertaining to forecasts only provide further indications that directional errors are particularly frequent for highly volatile series that show large changes which can vary in sign. This is so for the net change in business inventories and, very markedly, for

TABLE 21

Directional Errors in Selected Annual Forecasts and Extrapolations of GNP and Components and Industrial Production, 1953-63

(per cent)

Line	Predicted Variable	Forecast Sets and Periods Covered[a]			
		B 1953-63(11) (1)	F 1953-63(11) (2)	A 1958-63(6)[b] (3)	C 1958-63(6) (4)
1.	Gross national product	9.1	0	10.0[c]	16.7
2.	Consumption expenditures	0	9.1	0	0
3.	Gross private domestic investment	27.3	9.1	33.3	0
4.	Plant and equipment outlays	30.0[d]	n.a.	n.a.	0
5.	Producer's durable equipment	n.a.	11.1[e]	16.7	n.a.
6.	New construction	n.a.	50.0[f]	0	n.a.
7.	Inventory change	14.3[g]	30.0[d]	20.0[h]	16.7
8.	Net foreign balance	20.0[h]	30.0[d]	50.0	33.3
9.	Government expenditures	9.1	0	0	0
10.	Industrial production index	n.a.	9.1	10.0	16.7

(continued)

TABLE 21 *(concluded)*

Line	Predicted Variable	Extrapolative Models[i] N2* 1953-63(11)	N3 (1-2 terms) 1953-63(11)	N3 (2-5 terms) 1953-63(11)	N3 (range)[j] 1958-63(6)
11.	Gross national product	18.2	18.2	27.3	16.7 - 33.3
12.	Consumption expenditures	0	0	0	0
13.	Gross private domestic investment	27.3	27.3	36.4	33.3 - 50.0
14.	Plant and equipment outlays[d]	20.0	22.2	10.0	0
15.	Industrial production index	18.2	36.4	54.5	33.3 - 66.7

[a]Numbers of observations are given in parentheses following the specification of periods covered. Cases in which the numbers differ are identified in notes c - i.

[b]These refer to the last quarter of the next year, not to the total for the year (except for 1958).

[c]Refers to the period 1954-63 (10 years). The figure for 1958-63 is 16.3 per cent.

[d]Refers to the period 1953-62 (10 years).

[e]Based on nine observations.

[f]Based on eight observations.

[g]Based on seven observations.

[h]Based on five observations.

[i]Include the average-change projections N2* and autoregressive extrapolations N3 based on 1-5 lagged terms (see text for explanation of these models). The entries in column 2 refer to N3 with 1 or 2 terms (i.e., to extrapolations of the relationship between A_t and A_{t-i} , where $i = 1, 2$) and the entries in column 3 refer to N3 with 2-5 terms ($i = 2 \ldots 5$). The results in columns 1-3 cover the period 1953-63 and are therefore comparable to those shown for the forecast sets B and F in lines 1-4 and 10, columns 1 and 2.

[j]These figures refer to the period 1958-63 and are comparable with those shown for forecast sets A and C in lines 1-4 and 10, columns 3 and 4. Where two entries are given, the first one pertains to N3 models with fewer terms ($i = 1, 2, 3$) and the second one to models with more terms ($i = 2 \ldots 5$). Elsewhere (lines 12 and 14), the same result was obtained for each of the five N3 models used.

net foreign investment (lines 7 and 8). On the other hand, forecasts of government expenditures, like those of consumption, show very few or no directional errors (line 9).[27]

[27] The statements in this and the preceding paragraph are based on considerably more evidence than the selection presented in Table 21. In each case both forecasts and extrapolations were compared with the same set of "actual" data, which, as elsewhere, consist of the early vintage estimates available on an annual basis in the first quarter following the year to which the forecasts refer.