Volume Title: An Appraisal of Short-Term Economic Forecasts

Volume Author/Editor: Victor Zarnowitz

Volume Publisher: NBER

Volume ISBN: 0-87014-445-6

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

Publication Date: 1967

Chapter Title: Summary Measures and Types of Error

Chapter Author: Victor Zarnowitz

Chapter URL: http://www.nber.org/chapters/c0939

Chapter pages in book: (p. 31 - 59)

4

SUMMARY MEASURES AND TYPES OF ERROR

AVERAGE ERRORS OF ANNUAL FORECASTS OF GNP AND INDUSTRIAL PRODUCTION

The over-all averages in Tables 1 and 2 indicate that the errors in the annual forecasts of gross national product averaged about $9 to $10 billion disregarding sign. They appear small—about 2 per cent—compared with the average level of GNP, but they are big enough to make the difference between a good and a bad business year. The average year-to-year change in GNP over 1953–63 amounted to approximately $22 billion. Thus the errors were, in terms of absolute averages, not quite one-half the size of the errors that would have been produced by assuming that next year's GNP would be the same as last year's.

On the other hand, it is important to realize that aggregates as comprehensive and complex as GNP are beset by conceptual and estimation difficulties that make their measurement quite difficult and imprecise. When the series to be predicted is subject to substantial errors of observation, forecasting becomes particularly hazardous. A mean forecasting error of $10.0 billion does not appear very large relative to a mean of $5.5 billion in the revisions of the given series (see the absolute averages in Table 1, column 13).

In the index of industrial production, short-term fluctuations play a greater role and trend a lesser one than in GNP, which should make forecasting the index more difficult. On the other hand, revisions are much less disturbing for industrial production than for GNP. The relation between the forecasting errors and the variability of the data is not very different for the two series. The average errors of the annual forecasts of industrial production varied in the narrow range of 4 to 5 index points (1947–49 = 100). This compares with year-to-year changes averaging about 8 to 10 index points, as shown in Table 4 (columns 1 and 2, lines 16–22).

Table 4 collects all the salient statistics for an evaluation of the average performance of both GNP and industrial production forecasts in recent years. In order to facilitate comparisons, the measures for GNP are computed not only in billion dollars but also in index numbers (1947–49 = 100) to make them dimensionally similar to the figures for the industrial production index. Comparisons for the same sources and periods suggest that in most cases predicted changes in industrial production were somewhat larger than those in GNP, even though the actual changes were somewhat smaller in the former series (see columns 1 and 2, lines 9–22). The absolute errors tend to be larger for the industrial production forecasts (columns 3 and 4).

However, comparisons of this sort are difficult and necessarily crude when limited to absolute error measures. The main impression one gets here, for example, is that the changes and errors are quite similar for the two variables when expressed in index numbers with a common base. This is clearly insufficient, though interesting; it is necessary to move beyond such impressions, and to this end measures of relative forecast accuracy will be needed.

Measuring the errors relative to the level of the series, however, is not a satisfactory solution. Thus, a typical error of 2 per cent may be viewed as small for a series whose variations average 10 per cent, but it must be judged large indeed for a series whose variations average 1 per cent. To anticipate a theme developed later in this study, measures of relative accuracy should take into account the properties of the predicted variables which make for differences in the degree of difficulty that confronts the forecaster. An attempt to develop such measures will involve comparisons of forecasts with types of extrapolation. Meanwhile, there is still much to be learned from the absolute accuracy measures now under consideration.

ERRORS OF LEVELS, CHANGES, AND BASE VALUES

The simplest measure of error is obtained by comparing the predicted and the actual *levels,* but it is perhaps more important to compare predicted and actual *changes.* The error in the change will be the same as that in the level when the actual level at the time of the forecast is known. As a rule, however, it is not known, and the two will differ

TABLE 4

Annual Forecasts of GNP and Industrial Production: Average Actual and Predicted Changes and Average Errors of Forecasts of Levels, Changes, and Base Values, 1951-63

Line	Forecast Set[a]	Period Covered	Mean Absolute Changes		Mean Absolute Error, $\overline{\lvert E \rvert}$			Mean Error, $\bar{E}$		
			Actual[b] (1)	Predicted (2)	Level (3)	Change (4)	ECP[c] (5)	Level (6)	Change (7)	ECP[c] (8)
			Gross National Product (billion dollars)							
1.	A	1954-63	22.0	19.6	10.9	9.9	3.5	-7.2	-4.7	-2.4
2.	B	1953-63	21.7	19.1	9.7	8.4	2.5	-2.8	-1.3	-1.5
3.	C	1958-63	24.0	22.5	10.2	9.4	3.3	-3.2	-0.6	-2.6
4.	D	1956-63	23.1	18.3	9.4	8.1	2.7	-6.9	-4.2	-2.7
5.	E	1953-63	21.7	13.1	14.4	12.6	4.0	-10.8	-8.5	-2.3
6.	F	1953-63	21.7	20.2	8.1	6.3	2.9	-4.8	-2.3	-2.5
7.	G	1953-63	21.7	23.5	6.9	7.1	1.7	1.8	3.0	-1.2
8.	H	1954-63	22.0	18.5	10.4	9.5	3.1	-7.8	-5.8	-1.9
			Gross National Product (1947-49 = 100)							
9.	A	1954-63	8.9	7.9	4.4	4.0	1.4	-2.9	-1.9	-1.0
10.	C	1958-63	9.7	9.1	4.1	3.8	1.3	-1.3	-0.3	-1.0
11.	D	1956-63	9.4	7.4	3.8	3.3	1.1	-2.8	-1.8	-1.1
12.	E	1953-63	8.8	5.4	5.8	5.0	1.6	-4.4	-3.4	-0.9
13.	F	1953-63	8.8	8.2	3.3	2.6	1.2	-2.0	-1.0	-1.0
14.	G	1953-63	8.8	9.5	2.8	2.9	0.7	0.7	1.2	-0.5
15.	H	1954-63	8.9	7.5	4.2	3.8	1.2	-3.2	-2.4	-0.8

(continued)

TABLE 4 *(concluded)*

Line	Forecast Set[a]	Period Covered	Mean Absolute Changes		Mean Absolute Error, $\overline{\|E\|}$			Mean Error, $\bar{E}$		
			Actual[b] (1)	Predicted (2)	Level (3)	Change (4)	ECP[c] (5)	Level (6)	Change (7)	ECP[c] (8)
			Industrial Production (1947-49=100)							
16.	A	1954-63	8.6	8.4	4.8	5.0	0.6	-2.2	-2.0	-0.2
17.	C	1958-63	9.8	8.8	4.6	4.4	0.5	-0.3	-0.3	0.02
18.	D	1954-63[d]	8.6	7.2	4.5	4.1	1.0	-0.6	-1.0	0.4
19.	E	1951-63[d]	8.1	4.0	4.5	4.9	2.4	-0.4	-2.3	1.9
20.	F	1953-63	8.6	9.1	3.9	2.8	1.8	0.8	-0.7	1.5
21.	G	1953-63	8.6	9.3	4.2	3.6	0.9	2.5	1.9	0.6
22.	H	1954-63	8.6	7.6	3.8	4.1	1.0	-0.8	-1.1	0.4

[a]For a brief description of the forecasts, see Chapter 2 and Table 1, notes c and d.

[b]First estimates by the Department of Commerce (see Table 1, note a).

[c]Estimated current position, or base of the forecast. The base values are forecasters' own estimates for sets A (since 1957), B, C, F, and G. For the other sets, they are computed as follows: A (before 1957) – weighted averages of the last known position and the last reported relative change; D and H – sum of the last known level and the average annual change of the series up to the time of the forecast (N2* projection); E – weighted average of the last known level and the projection just described.

[d]Figures for industrial production forecasts covering periods comparable to GNP forecasts are:

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D	1956-63	7.4	8.0	3.9	4.0	0.6	-0.4	-0.5	+0.1
E	1953-63	3.7	8.6	4.8	5.0	2.5	-0.8	-2.8	+2.0

FORECASTS OF MAJOR EXPENDITURE COMPONENTS OF GNP

Forecasts of GNP are frequently built upon forecasts of GNP components, which are first derived separately and then combined, with varying amounts of attention given to their interdependence and consistency. Forecasts of total GNP may benefit from a partial cancellation of errors in forecasts of the components. This is definitely preferable to the opposite case of positively correlated and mutually reinforcing errors, which would make the predictions of GNP worse than those of its components and possibly invalidate them altogether. But gross inaccuracies in the component forecasts are, of course, always disturbing, even if these errors happen to be largely compensating. In extreme cases of this sort, the comprehensive aggregate forecast could be regarded as "good for the wrong reasons."

On the other hand, some methods of forecasting are concerned directly with measures of aggregate economic activity such as GNP, rather than with any GNP components or sectors of the economy. Forecasts using money supply, for example, fall into this category, as do indeed forecasts based on composite evaluation of business cycle indicators. These methods, therefore, may well yield better forecasts for GNP than for the components, which is quite understandable.

Annual forecasts of the major expenditure components of GNP have been analyzed for the following periods and sets: 1953–63, B and F; 1958–63, A and C.[4] Table 5 summarizes the results. It is based on errors computed by taking differences between the predicted and the actual percentage changes. The dollar levels of the GNP components differ drastically; the use of percentage changes enables us to make some comparisons between these variables that could not sensibly be made in terms of dollar changes.[5]

Looking at the summary measures in Table 5, column 1, one finds

[4] Forecasts A refer to the last quarter of the next year (at annual rates), not to the total for the year (except for 1958). They have therefore longer spans than the other forecasts which are all annual, and are not to be compared with the latter (forecasts with longer spans tend to have larger errors, see Chapter 5). This obviously does not affect the comparisons between forecasts for different variables (from any given source), which are the main concern of the present analysis.

[5] Using the symbols introduced earlier (see footnote 1), the error in predicting percentage change in a series from the base period t over the span i is defined as $\left(\frac{P_{\Delta(t+i)}}{A_t^*} - \frac{\Delta A_{t+i}}{A_t}\right) \cdot 100$, where $*$ denotes a preliminary estimate.

by the amount of error in the estimated current position or base of the forecast.[1]

Regrettably, the estimated current position (ECP) is not always reported by the forecasters. Where such base estimates are not given they must be imputed, if the change forecasts are to be analyzed. Our imputations are based on extrapolative methods; where reported ECP's are given in some years but not in others, the estimates for the latter also utilize the information for the former.[2] The imputations cannot be shown to be "wrong" (or "right") but undoubtedly have shortcomings. All that can be claimed is that they seem on the whole reasonable when compared with the reported ECP's.

The ECP errors are, of course, typically smaller than the errors in the forecasts proper. One would expect the present or the recent past to be better known than the future. Looking at the mean absolute errors in Table 4, one finds no exceptions to this rule (compare columns 3 and 5). However, the ECP errors are by no means negligible. They average about one-fourth or one-third of the corresponding errors of the level forecasts for GNP, and from an eighth to more than half for industrial production.

The error of each level forecast is the algebraic sum of the ECP error and the error of the predicted change (see footnote 1). For the GNP forecasts, errors of base have on the average the same (negative) signs as the larger errors of the future levels (Table 4, columns 6 and 8). Accordingly, the errors of change tend to be less than the errors of level (compare columns 3–4 and 6–7).[3] For industrial production, errors of base and of level often differ in sign, and there is less regularity in the relation between the level and the change forecasts (Table 4, lines 16–22).

[1] Let the level error be defined as $E_{t+1} = P_{t+1} - A_{t+1}$, where P, A, and E denote the prediction, the actual value, and the error, respectively; t is the current and $(t + 1)$ the next year. The predicted change is $P_{\Delta(t+1)} = P_{t+1} - A_t^*$ where A_t^* is a preliminary estimate of the current position (ECP). The error of the change forecast is $E_{\Delta(t+1)} = P_{\Delta(t+1)} - \Delta A_{t+1}$, where $\Delta A_{t+1} = A_{t+1} - A_t$. The error of ECP is $E_t = A_t^* - A_t$.

It follows directly from these definitions that

$$E_{\Delta(t+1)} = P_{t+1} - A_t^* - (A_{t+1} - A_t) = E_{t+1} - E_t.$$

Hence, if $A_t^* = A_t$ (the ECP error is zero), then $E_{\Delta(t+1)} = P_{t+1} - A_{t+1} = E_{t+1}$.

[2] For brief descriptions of how the ECP's were computed, see Table 3, note c.

[3] The errors of the annual level forecasts for GNP average \$10 billion when taken without regard to sign; the errors of the corresponding change forecasts average slightly more than \$8 billion, or 18 per cent less (see Tables 1 and 2, line 12, col. 13).

TABLE 5

Forecasts of Relative Changes in Major Components of GNP: Summary Measures of Error, 1953-63

(percentage points)

Line	Predicted Variable	Mean Absolute Error[a] $\overline{\lvert E \rvert}$ (1)	Mean Error[a] $\overline{E}$ (2)
	Forecast Set B: 1953-63		
1.	Gross national product	1.88	-0.31
2.	Personal consumption expenditures	1.25	-0.16
3.	Gross private domestic investment	11.69	-2.23
4.	Plant and equipment	6.42	-3.51
5.	Housing	10.95	-6.83
6.	Total government expenditures	1.99	+0.81
7.	Federal	4.29	+0.91
8.	State and local	1.46	-1.24
	Forecast Set F: 1953-63		
9.	Gross national product	1.44	-0.60
10.	Personal consumption expenditures	1.51	-0.70
11.	Gross private domestic investment	6.20	-1.42
12.	Total government expenditures	1.19	+0.42
	Forecast Set F: Other Periods[b]		
13.	Consumer durables	4.44	+2.09
14.	Consumer nondurables	1.24	+0.92
15.	Consumer services	1.64	-0.88
16.	Producers' durables	5.01	-1.79
17.	Nonresidential construction	3.00	-1.02
18.	Residential nonfarm construction	7.00	-3.83
19.	Federal government expenditures	1.37	+0.10
20.	State and local expenditures	1.05	-0.76
	Forecast Set A: 1958-63[c]		
21.	Gross national product	1.79	+0.02
22.	Personal consumption expenditures	0.92	+0.05
23.	Consumer durables	4.57	+0.67
24.	Gross private domestic investment	9.08	+2.45
25.	Producers' durables	5.50	-2.70
26.	New construction	3.85	-3.78
27.	Total government expenditures	2.73	-0.83

(continued)

TABLE 5 *(concluded)*

Line	Predicted Variable	Mean Absolute Error[a] $\overline{\lvert E \rvert}$ (1)	Mean Error[a] $\bar{E}$ (2)
	Forecast Set C: 1958-63		
28.	Gross national product	1.87	-0.07
29.	Personal consumption expenditures	1.36	-0.25
30.	Gross private domestic investment	8.03	+1.83
31.	Plant and equipment	4.72	+0.11
32.	Residential construction	6.97	-1.90
33.	Total government expenditures	1.44	-1.04
34.	Federal	2.92	-0.89
35.	State and local	1.90	-1.90

[a]Based on errors of percentage change as defined in text and footnote 5.

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

[c]Refer to the last quarter of the next year (at annual rates), not to the total for the year (except for 1958, which is an annual forecast). See note 4.

that the errors in predicting percentage changes in personal consumption are far smaller than those in forecasts of gross private domestic investment (GPDI). The errors for total government spending are moderate: larger than those in consumption but much smaller than those in investment.

Within aggregate consumption, the errors are much larger for durable goods than for either nondurables or services. Within aggregate investment, the errors for housing or residential construction are particularly large: they exceed the errors for plant and equipment (in sets B and C) and the errors in both producers' durables and nonresidential construction (in set F). A different breakdown shows the record for producers' durable equipment to be worse than that for total new construction (in set A).[6]

[6] Taken together, these results suggest that forecasts of nonresidential construction must be considerably better than those of producers' durables, offsetting the very large housing errors. This is actually so in set F for which a construction breakdown is available (see lines 16–18 in the table).

These results agree with what would be generally expected. Spending for nondurables and services is dominated by stable growth tendencies, which lend a measure of stability to over-all consumption, while outlays for consumer durables are much more volatile. Investment is, needless to say, much more volatile than consumption. Expenditures on producers' durables have been less stable than construction in recent times. The more volatile a series, the more difficult is the prediction of its relative changes and the greater the probable forecast error.

Expenditures on housing have been restrained during business expansions by a scarcity of mortgage credit and stimulated during recessions by greater availability of credit. These countercyclical effects cause spending on residential construction to behave quite differently from other major expenditure categories. Failure to recognize such differences may be largely responsible for the particularly bad showing of the housing forecasts.

The errors in percentage change forecasts for total GNP are, on the average, much smaller than those for most of the component types of expenditures. Thus, only consumption and, in some forecasts, total government or state and local expenditures were predicted as well as, or better than, total GNP (Table 5, column 1). This indicates that the aggregation by sectors has, in fact, been associated with a very substantial cancellation of errors.

Measures of relative change errors need to be supplemented by measures in dollar terms for two reasons. First, the mean arithmetic errors of the sectoral forecasts add up to the average error for total GNP, enabling us to observe the extent to which errors for the different types of expenditure either cumulate or offset each other and also to compare the corresponding averages without regard to the signs of the errors. Second, the components of GNP that become negative—net change in inventories and net foreign investment—can be analyzed only in absolute, not in relative, terms.

Table 6 summarizes the errors of absolute changes for the two main sources of sectoral forecasts in our annual data—sets B and F. The means of these errors, like those in Table 5, column 2, are predominantly negative (columns 2 and 4), which is typical for predictions of changes in total GNP as well as industrial production (see Table 4, col. 7). Only government expenditures and net changes in inventories

TABLE 6

Summary Measures of Error in Two Sets of Forecasts of Absolute Changes in Major Components of GNP and Investment, 1953-63

(billion dollars)

Line	Component of GNP or Investment	Forecast Set B: Mean Absolute Error, $\overline{\|E\|}$ (1)	Forecast Set B: Mean Error, $\bar{E}$ (2)	Forecast Set F: Mean Absolute Error, $\overline{\|E\|}$ (3)	Forecast Set F: Mean Error, $\bar{E}$ (4)
1.	Plant and equipment	2.64	-1.29		
2.	Producers' durables			1.21[a]	-0.41[a]
3.	Residential construction	1.59	-0.92	1.29[b]	-0.79[b]
4.	Nonresidential construction			.56[b]	-0.19[b]
5.	Net change in inventories	2.75	+0.76	2.16[c]	+1.10[c]
6.	Gross private domestic investment	5.86	-1.43[d]	3.87	-0.60
7.	Personal consumption expenditures	3.79	-0.41	4.16	-1.67
8.	Government expenditures	1.76	+0.80	1.06	+0.33
9.	Net foreign investment	1.35	-0.29	1.36	-0.37
10.	Gross national product	8.37	-1.28[e]	6.34	-2.33[e]

[a] For 1955-63.

[b] For 1956-63.

[c] For 1954-63.

[d] Sum of lines 1, 3, and 5 does not add to total in line 6 because of rounding.

[e] Sum of lines 6-9 does not add to total in line 10 because of rounding.

have positive mean errors in both sets. The averages for the component figures are, in general, smaller absolutely than the corresponding figures for GNP (the investment forecasts in set B provide an exception). They conceal a greal deal of variation in sign among the errors, as witnessed by the large differences in absolute size between the paired entries in columns 1–2 and 3–4 (this cancellation of errors is particularly pronounced in forecasts of set B).

BIAS IN FORECASTING

A set of forecasts is said to contain a bias if it typically understates or overstates the corresponding actual values. Thus, the criterion of the absence of bias would be that in the "long run" (that is, for sufficiently large numbers of comparable forecasts), predictions and realizations should be on the average equal.[7] Actually, strict equality of the averages cannot be expected in the limited samples that can be observed, and the criterion is to be understood to mean that for unbiased forecasts the difference between the averages is not significant in the statistical sense.

To give some examples of biased forecasts, consider a fluctuating series which rises in good times and falls in bad ones. A strong pessimistic bias would be illustrated by forecasts that consistently specified lower levels than those observed for that series, which implies predictions of too small increases and too large decreases. A strong optimistic bias is the reverse: forecasts of higher than the actual levels, that is, of too large increases and too small decreases.

A simple measure of bias is the mean error, i.e., the difference between the means of actual and predicted values. Each observed error can thus be thought to consist of a bias, which is the average error over the entire period, and the deviation of the observed error from the average. The bias, therefore, is the constant element in the errors, since it is the same for each observed error, while the remainder reflects the variation among the recorded errors measured from this average.

A measure of the over-all accuracy of forecasts, which can be conveniently used to separate the bias from the remaining error, is the mean square error $M_P{}^2$, computed by squaring the individual forecast errors and averaging the results. The mean square error is the sum of two components: the square of the mean error (or bias), and the variance of the errors.[8]

The mean square error analysis involves more technical apparatus and language than I wish to use in this paper; the interested reader may consult another report in the National Bureau project on forecasting for a further discussion of the statistical concepts just outlined

[7] This is just a translation into a less technical language of the standard statistical definition of bias as the inequality of "expected values."

[8] For n time periods (t),

$$M_P{}^2 = \frac{1}{n} \sum_{t}^{n} (P_t - A_t)^2 = \bar{E}^2 + S_E{}^2,$$

where $\bar{E}^2 = (\bar{P} - \bar{A})^2$ is the squared mean error and $S_E{}^2$ is the variance of errors. In the absence of bias, $\bar{P} = \bar{A}$ and $M_P{}^2 = S_E{}^2$.

and a presentation of some corresponding estimates.[9] This section will merely summarize some of the results of this analysis and rely mainly on the simpler measures given in the tables of the present report.

The mean errors of the annual forecasts in our collection are given in Tables 4, 5, and 6 above. They vary greatly in size relative to the average magnitude of the changes that the forecasters tried to predict (compare columns 1 and 7 in Table 4) and also relative to the corresponding mean absolute errors. Tests of the statistical significance of the mean errors are presented elsewhere (see the reference in note 9). They confirm that the relative importance of bias varies substantially among forecasts from different sources and for different variables. Thus the tests give considerable evidence of bias in some of the GNP forecasts, but very little indication of significant bias in the industrial production forecasts. The differences among forecasters are probably less meaningful; they cannot be ascribed entirely to the variation in ability or technique because the periods covered are not identical for all the sets.[10] In general, however, these tests lack conclusiveness, since the periods covered by our data are short and, hence, the samples of observations per forecast set are small.

A point of interest in the GNP forecasts is that they generally start from estimates of current levels that are too low. The mean errors in these estimates (ECP) are negative for all forecast sets (Table 4, column 8, lines 1–8). This is presumably due in part to the forecasters' reliance on early figures that are often revised upward during the year to which the forecasts refer.[11] The base values for industrial production, a variable for which such revisions are less frequent and less important, seem to be in most cases somewhat overestimated (Table 4, column 8, lines 16–22). In any event, the base estimates are often substantially biased and, consistent with this, the elements of bias tend to be larger in the level forecasts than in the corresponding change forecasts.

[9] Jacob Mincer and Victor Zarnowitz, "The Evaluation of Economic Forecasts" (forthcoming).

[10] For example, the forecast set A for GNP shows a strong bias, with the squared mean error $\bar{E}^2$ accounting for approximately 40 per cent of the total M_p^2 (see note 8 for the formula used). But this feature of set A is due mainly to large underestimation errors in the early postwar years (see first section in Chapter 3).

[11] Possibly some forecasters are aware of this bias in their estimates and try to compensate for it implicitly in their change predictions rather than explicitly by correcting their ECP figures. This assumption would be consistent with one of our forecast sets, in which the base levels are underestimated but the changes in GNP are overestimated (Table 4, line 7).

Other studies have suggested that forecasters generally underestimate both rises and declines in the predicted series.[12] In this case, the *absolute* average and the variance of the predicted changes would be smaller than those of the actual changes, but there need be no bias, in the specific sense of a significant difference between the arithmetic averages of predictions and realizations. Indeed, underestimation of changes would not constitute a systematic error that forecasters could or should guard against if it were merely the result of forecasters failing to predict random variations in the actual values. A forecaster who ignored only such variations and succeeded in predicting all other changes would have done as well as could be hoped for, yet his forecast would have a variance smaller than that of the actual values and, in this sense, would necessarily "underestimate" the observed changes.[13]

As this illustrates, it is important to recognize that forecasting errors which can be traced directly to short random movements must be regarded as unavoidable. Thus, a sudden outbreak of war or a strike started without warning are events that an economic forecaster can hardly be expected to predict (though his job certainly does include an evaluation of the effects of such events, once known, on the economy). To put it differently, in principle the requirement of a good forecast is that it predict well the systematic movements of the given variable, not that it predict the actual values, since random elements are virtually always present in economic time series.

Underestimation of changes would have a different meaning if the changes pertained to longer cyclical movements, not just to short irregular variations. This result could come about if forecast errors varied systematically with the values predicted so as to yield underestimates at high and overestimates at low levels. Predictions with this property would have a larger over-all error than predictions which are independent of levels, hence elimination of this type of systematic

[12] Franco Modigliani and Owen H. Sauerlander, "Economic Expectations and Plans of Firms in Relation to Short-Term Forecasting," in *Short-Term Economic Forecasting*, Studies in Income and Wealth 17, Princeton for NBER, 1955, Table 8 (based on the *Fortune* and Dun & Bradstreet Surveys), pp. 288–289; Henri Theil, *Economic Forecasts and Policy*, Amsterdam, 1958, Chapters III–V.

[13] Consider the equation $A_t = a + bP_t + u_t$. Unbiased, efficient forecasts require that $a = 0$, $b = 1$, and that u_t be a random, nonautocorrelated variable with mean zero. In this case, the variance of A will equal the sum of the variances of P and of u.

underestimation error must be viewed as desirable.[14] But again, errors of this kind do not necessarily involve a bias, that is, a discrepancy between long-run averages computed with regard to sign.[15] This can be perceived intuitively by visualizing forecasts that result in a series with cyclical movements around a trend smaller than those actually recorded, but with a correct estimate of the trend itself. Here, then, is a type of systematic (and potentially serious) error which is not encompassed by the usual statistical definition of bias.[16]

Have the forecasters represented in our collection tended to underestimate changes of either the random or the systematic kind, or both? The hypothesis that forecasters manage to reproduce the time path of the predicted series in some smoothed form receives little support from the materials under review. Actual changes are in fact usually reduced in the forecasts, but not regularly in both the upward and downward direction as the hypothesis would imply.[17] The forecast errors are by no means limited to the random components of the series, though such irregular movements doubtless account for a large part of the fore-

[14] If P_t and E_t are correlated, $b \neq 1$ in the relation $A_t = a + bP_t + u_t$ (in the case of underestimation of systematic changes, $b > 1$). The adjusted forecast $P_t^* = a + bP_t$ would be an improvement on the original forecast P_t, whenever the expected values of a and b deviate from zero and unity, respectively. Thus, if $b \neq 1$ then the variance of E_t is larger than that of u_t (i.e., $S_E^2 > S_u^2$), since $S_E^2 = S_u^2 + (1 - b)^2 S_P^2$ (where S_P^2 is the variance of P_t).

[15] From the formulas for M_P^2 and S_E^2 in notes 8 and 14, it follows that the mean square error can be decomposed into three parts:

$$M_P^2 = \bar{E}^2 + S_E^2 = (\bar{P} - \bar{A})^2 + (1 - b)S_P^2 + S_u^2.$$

In unbiased forecasts, $\bar{P} = \bar{A}$ and the first component of M_P^2 is zero; but the second component will be positive, unless $b = 1$ (it is an increasing function of the deviations of b from zero). It may be noted that this decomposition is equivalent to one of the two forms introduced in Henri Theil, *op. cit.*, pp. 34–39.

[16] One must remember, however, that bias is a possible property of sample estimates referring to some population or aggregate of phenomena to be studied; hence, what bias is depends on the definition of that aggregate. If a good reason existed to treat different ranges of the series as belonging to different "populations," then bias in the strict sense could no longer be excluded in the above situation. For then the ranges would in effect be viewed as separate variables, and underestimation of one set of values (say, the high levels) would constitute one case of bias, while overestimation of the other set (low levels) would constitute another. In each range, predictions would differ on the average from realizations both for levels and changes.

[17] Presumably, smoothing out the irregular movements would work in both directions alike and not affect systematically the mean of the changes over time (this assumes that such movements themselves have a zero mean).

casters' difficulties.[18] As was shown in Chapter 3, forecast errors tend to differ depending on cyclical phase, with large underestimates being concentrated mainly in early expansion periods, which are typically periods of high growth rates.

Actually, it was primarily the increases in GNP that were underestimated in the forecasts reviewed, not the decreases (as will be demonstrated in the section that follows). Underestimation of increases alone would be sufficient to produce the observed result that the predicted *levels* of GNP, which is a series with an upward trend, were on the average too low (note the predominance of negative signs in column 6 of Table 4).[19] The same applies to the finding that, in terms of averages taken without regard to sign, actual changes exceeded predicted changes for twelve of the fifteen forecast sets recorded in Table 4 (columns 1–2). It is only when increases and decreases are treated separately that one can obtain meaningful indications of a downward bias in the forecasts—an underestimation of growth.

UNDERESTIMATION OF GROWTH

Forecasters often regard themselves as "conservative" or "cautious." If this means cautious in appraising growth prospects, the results of our analysis bear out this view.

Chart 4 presents scatter diagrams for selected sets of annual forecasts, which relate the actual to the predicted changes. The chart is followed by a key, which is self-explanatory.

Because of the prevalence of upward trends in such series as GNP, personal consumption, and industrial production, the bulk of the observations in the corresponding scatters fall into the first quadrant

[18] The proportion S_u^2/M_p^2 in the GNP and industrial production forecasts (see note 15 for the underlying decomposition of M_p^2) is very seldom less than 0.5 and often as high as 0.9 or more (in which case, the other "systematic" components of M_p^2 are actually unlikely to be significant). However, it should be noted that the u's are here simply the residuals from the regression of A_t on P_t, i.e., that part of A_t which the "predictor" P_t was unable to account for; they need not be in fact purely random and nonautocorrelated.

[19] When increases are underestimated, the future levels of the series will also as a rule be underestimated. (This follows necessarily if the current position is known or itself understated.) On the other hand, underestimation of *decreases* would tend to result in *over*estimation of levels. Since rises are more frequent than declines in the series considered here, errors relating to rises, if sufficiently systematic and large, are likely to dominate the over-all result.

CHART 4

Scatter Diagrams of Relations Between Predicted and Actual Changes, Selected Annual Forecasts of Four Aggregative Variables, 1953–63

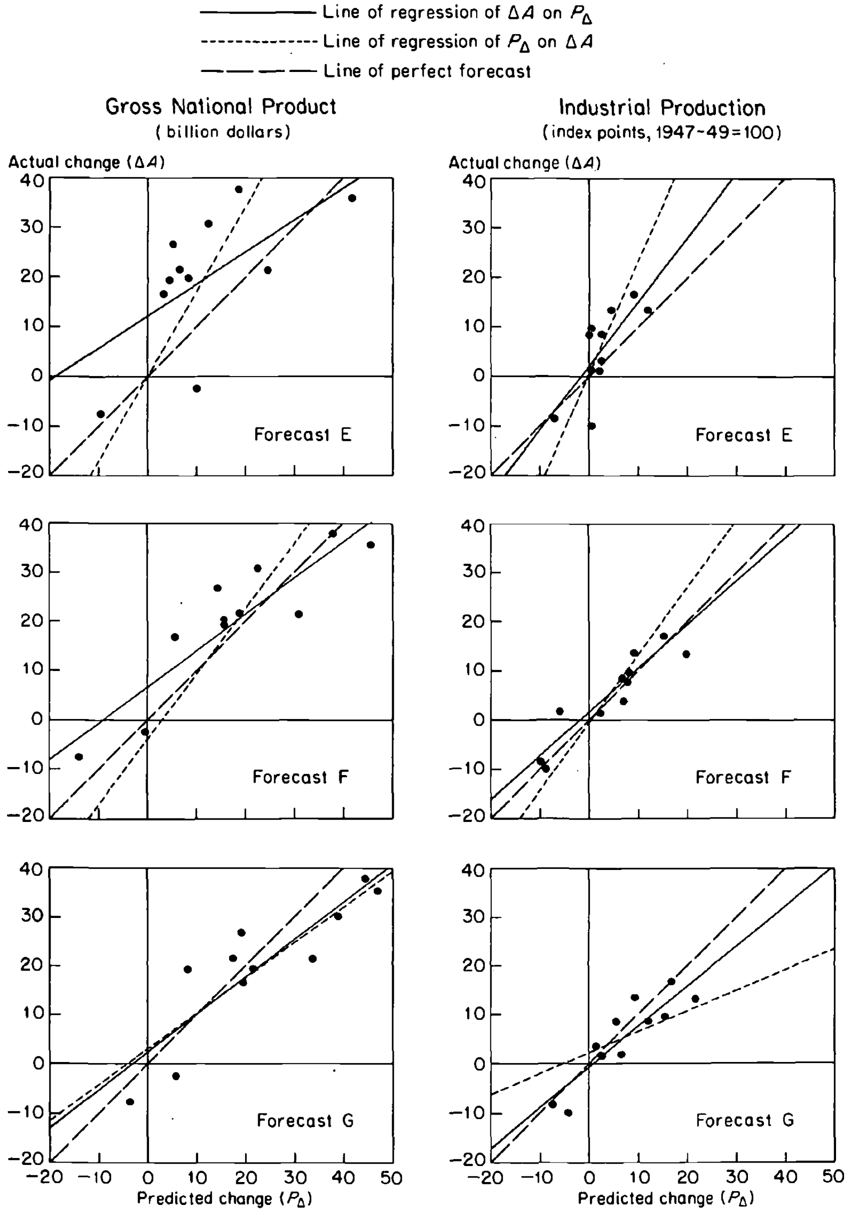
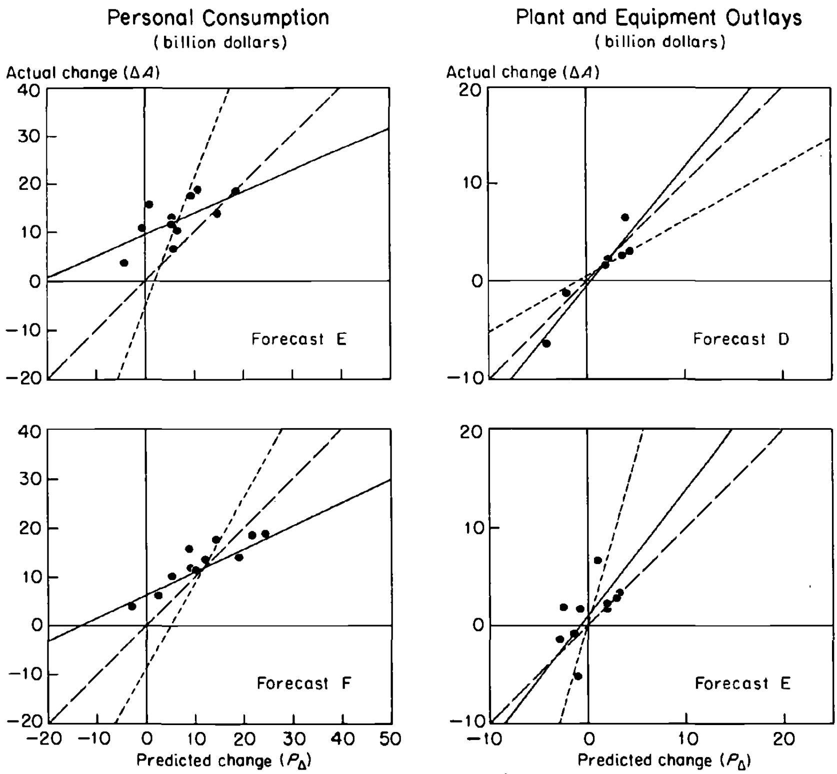
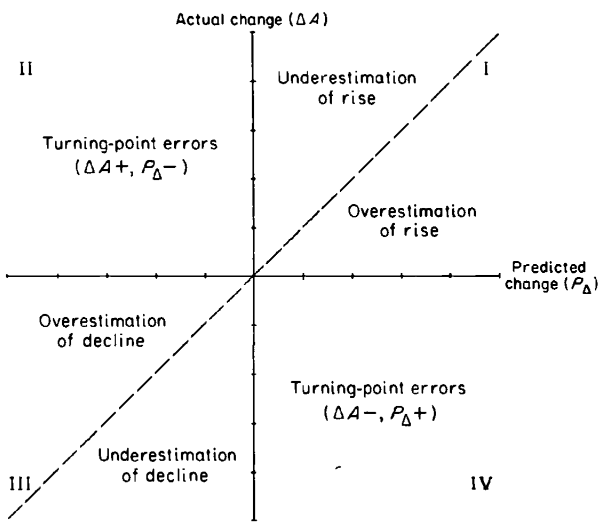


CHART 4 (*concluded*)



KEY



(where the actual and predicted changes, ΔA and P_{Δ}, are both positive). That is, few declines occurred in the annual values of GNP and few were expected, hence there are relatively few points in the other quadrants. Also, there are often more points above than below the 45° line of perfect forecasts in quadrant I. This means that underestimates of increases are more frequent than overestimates.

Reflecting these characteristics of the scatters, the regression lines in Chart 4 run, in large part, through the areas of underestimation of changes.[20] This, however, is by no means always true, as illustrated in the diagrams for forecasts G where the overestimates carry more weight than the underestimates.

For series that are more volatile, such as the plant and equipment outlays, the scatters show no distinct concentrations of points in any single quadrant. Increases are less dominant here and less confidently predicted. Sometimes declines have been expected instead, giving rise to turning point errors in quadrant II; at other times, declines have been correctly predicted (quadrant III). The combination of predicted rises and actual declines is not recorded in our graph for the plant and equipment outlays (quadrant IV is empty there) but a few examples of such turning-point errors are found in other investment forecasts. As this illustration suggests, forecasts of such variables show little evidence of biases but also low correlations with the observed changes.

While the results for the different forecast sets vary in several respects, including the significance of the regression estimates, as would be expected, it is worth noting that in most cases the regressions of the actual on the predicted changes show positive intercepts and slopes less than one (see the solid regression lines in Chart 4).

Table 7 summarizes the distributions of errors in the forecasts of annual changes according to whether they represented overestimates, underestimates, or turning-point (directional) errors. It confirms that forecasters frequently underrated the increases in GNP, while displaying no such tendency in regard to decreases.

In interpreting this asymmetry, it is necessary to recall that the recent recessions in the United States have been sufficiently mild and short to cause only very small declines, or merely retardations of

[20] Note in particular the diagram for the industrial production forecast E, in which the scatter and both regression lines clearly indicate a tendency to underestimate increases and decreases.

TABLE 7

Forecasts of Annual Changes in Four Comprehensive Aggregates: Distribution by Type of Change and Type of Error, 1952-63

Line	Type of Change[a]	Number of Forecasts of Annual Changes: Total (1)	Under-estimates[b] (2)	Over-estimates[c] (3)	Turning-Point Errors[d] (4)	Probability of at Least as Many Under-estimates[e] (5)
		Gross National Product (8)[f]				
1.	All observations	78	46	25	7	.008
2.	Increases	64	43	21	0	.004
3.	Decreases	14	3	4	7	.756
		Personal Consumption Expenditures (4)[f]				
4.	All observations[g]	39	26	10	3	.006
		Gross Private Domestic Investment (3)[f]				
5.	All observations	28	12	12	4	.581
6.	Increases	18	8	7	3	.500
7.	Decreases	10	4	5	1	.746
		Industrial Production (7)[f]				
8.	All observations	70	37	26	7	.104
9.	Increases	57	28	23	6	.288
10.	Decreases	13	9	3	1	.073

[a]Increases and decreases refer to the direction of changes in the actual values (first estimates for the given series).

[b]Predicted change is less than actual change ($P_{\Delta(t+1)} < \Delta A_{t+1}$).

[c]Predicted change exceeds actual change ($P_{\Delta(t+1)} > \Delta A_{t+1}$).

[d]Sign $P_{\Delta(t+1)} \neq$ sign ΔA_{t+1}.

[e]Based on the proportion of all observations, other than those with turning-point errors, accounted for by the underestimates (i.e., col. 2 divided by the difference between col. 1 and col. 4). Probabilities taken from Harvard Computation Laboratory's *Tests of the Cumulative Binomial Probability Distribution,* Cambridge, Mass., 1955.

[f]The figure in parentheses is the number of forecast sets covered.

[g]All observed changes are increases.

growth, in the *annual* values of GNP. According to early estimates, declines occurred in 1953–54 and 1957–58.[21] On the former occasion most forecasts showed larger declines than actually occurred, and on the latter most of them missed the downturn. In the years when GNP increased, on the other hand, underestimates were almost twice as frequent as overestimates (see Table 7, lines 2 and 3). In forecasts of personal consumption expenditures, a series which has risen steadily in recent years, errors of underestimation also prevail (Table 7, line 4).

In contrast, overestimates are about as frequent as underestimates in forecasts of gross private investment (lines 5–7). In this series the trend is much weaker than in the others, and the cyclical and irregular movements are relatively stronger. Also, for industrial production, a series with more pronounced fluctuations than those in GNP in current dollars and with a weaker trend, the forecasts show less tendency to underestimate increases. On the other hand, the production index declined much more in recessions than GNP did, and its contractions were in fact very often understated (Table 7, lines 8–10).

In short, these results suggest that the notion that forecasters tend to underestimate changes in general—that is, increases as well as decreases, short fluctuations as well as longer trends—is a gross oversimplification. Important asymmetries seem to exist here between errors in predicting upward and downward changes and also between forecasts for series with strong growth trends and others.[22]

Examination of forecasts of other GNP components yields additional evidence that is consistent with the hypothesis of underestimation of growth. For all the major expenditure components taken together, increases were underestimated nearly half the time and overestimated more than one-third of the time, while turning-point errors account for the remaining one-sixth of these observations. Decreases were underestimated about as often as overestimated; they were also, on the whole, just as frequently missed, but the proportion of these turning-

[21] Subsequent revisions converted the declines of GNP in 1953–54 and 1957–58 into small *increases* (see Table 2, lines 1–2, columns 2 and 6).

[22] The last column of Table 7 lists the probabilities of observing at least as many underestimates as shown by the forecasts in question. These are derived under certain highly simplified assumptions which permit application of the binomial distribution: that the forecasts are independent and the chances of under- and overestimation are equal in all cases with no turning-point errors. These assumptions can be questioned and the tests cannot be interpreted in any strict sense, but the results are suggestive in the context of the preceding observations.

point errors varies greatly (set B missed half the declines, set F just one-sixth).

As this indicates, all types of error are well represented in these forecasts, which reflects in part the diverse behavior patterns of the series concerned. The prevalence of negative signs among the mean errors (Table 5, col. 2, and Table 6, columns 2 and 4) is due not simply to underestimation of changes but to a combination of factors: over 70 per cent of the observations relate to *increases* in the series, where underestimation is common; and most of the turning-point errors are negative, being false signals of declines.[23]

To conclude, asymmetries in the distribution by type of errors associated with increases and decreases cast doubt upon the simple hypothesis of general underestimation of changes, but there is considerable support for the idea that forecasters tend to underrate increases or growth of comprehensive economic aggregates. As a closely related point, there is also some direct evidence that underestimation occurs primarily in certain phases of business cycles, particularly in the early upswing (see Table 3 and text above). Such patterns probably would not be observed if this type of error applied merely, or even largely, to short erratic changes.

TURNING-POINT ERRORS

Many economic time series, and especially those of comprehensive coverage, show strong systematic movements of substantial duration—trends and specific cycles. They are positively autocorrelated, often to a high degree. These observations led to the now widely held idea that it should be rather easy to predict a continuation of the rise or fall in these series; to forecast correctly the end of the current movement or phase appears to mark a more meaningful predictive success.

To appraise the performance of forecasts with respect to turning-point errors, one must compare the signs of predicted and actual changes. A forecast for any future period is based on the estimate of the current value of the series, not on the past forecast that has already been superseded by the intervening information; hence forecasts

[23] Errors of change (defined as the forecast minus the actual figure) are negative when increases are underestimated or decreases are overestimated and when the predicted change is negative and the actual one is positive. In our samples, the totals of such cases are typically greater than those of the converse cases in which the errors of change are positive.

made at successive points of time do not form a continuous series. Inferences from level forecasts treated as a series can easily be wrong, particularly on turning-point errors.[24]

Suppose a series rises, declines, and rises again in three consecutive years, while forecasts indicate three increases. Looking at the levels, one may conclude that the forecaster missed two turning points, a peak and a trough. Looking at the patterns of change, which are + − + for A_t and + + + for P_t, one finds a single directional error, relating to the change between years 1 and 2. Actually, only one turning-point error was made, since the forecaster knew at the end of year 2 that a decline had taken place and decided that it would be reversed the following year. In short, to determine whether a turning point has been forecast, one must compare the forecast change with the preceding actual change, not with the preceding forecast change.

In evaluating turning-point errors, two questions must be asked: (1) How often do turning points occur that have not been predicted? (2) How often do predicted turns actually occur? Accordingly, there are two basic types of error: missed turns and false signals. In the preceding paragraph, an example of a missed turn was provided. A false signal would be given in the case of reversed patterns: + + + for A_t and + − + for P_t.[25]

In addition to the two types of error, there are two types of correct forecasts. One is when a turning point (TP) was predicted and it did

[24] Consider the following tabulation on the actual (A) and predicted (P) values of industrial production. It will be noted that the troughs (T) fall in the first half of 1958 for A and in the second half of 1958 for P—an apparent lag of the latter behind the former. But the reversal of sign from minus to plus occurred at the same time in both the actual and the predicted change, which indicates the absence of a real lag (the directional change is denoted by *). The important point here is that such a situation can readily be produced by underestimation of changes on both sides of the turning point.

Date (t)	A_t	$(A_t - A_{t-1})$	P_t	$(P_t - A_{t-1})$
2nd half 1957	142			
1st half 1958	129 (T)	−13	138	−4
2nd half 1958	138	+9*	130 (T)	+1*
1st Q 1959	145	+7	143	+5

[25] Reversing the order of presentation used above, one can say that type I error arises when a turning point was predicted but did not occur; type II error, when a turning point was not predicted but did occur. This parallels the familiar statistical dichotomy of first-kind and second-kind errors, the former consisting of an incorrect rejection, the latter of an incorrect acceptance of the null hypothesis (which is here represented by a forecast of no directional change).

occur; the other, when no TP was predicted and none occurred. Thus there are four basic possibilities which can be arranged in a 2×2 table as follows. (N refers to the absence, T to the presence of a turning point. The first letter refers to actual values, the second to forecasts.)

Forecast / Actual	No TP	TP
No TP	NN	NT
TP	TN	TT

The number of correct forecasts is the sum of the diagonal frequencies: $NN + TT$. False signals are represented by NT and missed turns by TN. To compute the proportion of the former, one should use as a base the number of all predicted turns. This yields the ratio $\bar{E}_{T1} = \frac{NT}{NT + TT}$. In the proportion of missed turns, on the other hand, the base is the number of all recorded turns $\bar{E}_{T2} = \frac{TN}{TN + TT}$. These proportions of turning-point errors are analogous to mean errors, hence the choice of symbols.[26]

If forecasters relied mainly on trend projections, their record would be relatively good on false warnings and poor on missed turns.

The tendency for a comprehensive aggregate such as GNP to grow most of the time is well known to economic observers; hence few reversals of direction from one year to another are recorded for this series, and also few are predicted (note the high proportions of NN in Table 8, column 2). False signals are particularly infrequent. Only two are listed for the annual GNP predictions, and these refer to the first two years covered, 1947 and 1948, when an early postwar slump was still widely anticipated (see column 7, line 1).

For variables with stronger cyclical and irregular components, turning-point predictions should, and probably will, be more numerous. Although the chances that a predicted turn will actually occur are

[26] Assign to each case in the category NT the value 1 and to each case in the category TT the value 0; the mean of these values is $\bar{E}_{T1}$ and their variance is $S^2_{T1} = \bar{E}_{T1}(1 - \bar{E}_{T1})$. Similarly, assign to each instance of TN the value 1 and to each instance of TT the value 0; here the mean of the resulting scores is $\bar{E}_{T2}$ and their variance is $S^2_{T2} = \bar{E}_{T2}(1 - \bar{E}_{T2})$.

TABLE 8

Frequency of Turning Points and Errors in Annual Forecasts of GNP and Industrial Production, 1947-63

			Number of Years[b]		Number of Turning Points[b]					Percentage[b] of	
Line	Fore-Cast Set[a]	Period Covered	Covered (NN+TT+TN+NT) (1)	With No TP Observed or Predicted (NN) (2)	Observed (TT+TN) (3)	Predicted (TT+NT) (4)	Correctly Predicted (TT) (5)	Missed (TN) (6)	Falsely Predicted (NT) (7)	Observed Turns Missed[c] ($\bar{E}_{T2}$) (8)	Predicted Turns False[d] ($\bar{E}_{T1}$) (9)
					Gross National Product						
1.	A	1947-49, 1954-63	13	8	3	4	2	1	2	33	50
2.	B	1953-63	11	9	2	1	1	1	0	50	0
3.	C	1958-63	6	5	1	0	0	1	0	100	0
4.	D	1956-63	8	7	1	0	0	1	0	100	0
5.	E	1953-63	11	9	2	1	1	1	0	50	0
6.	F	1953-63	11	9	2	2	2	0	0	0	0
7.	G	1953-63[e]	11	9	2	1	1	1	0	50	0
8.	H	1954-63	10	8	2	1	1	1	0	50	0
					Index of Industrial Production						
9.	A	1947-49, 1954-63	13	9	3	4	3	0	1	0	25
10.	C	1958-63	6	4	1	2	1	0	1	0	50
11.	D	1947-63	17	10	4	6	3	1	3	25	50
12.	E	1951-63	13	10	3	1	1	2	0	67	0
13.	F	1953-63	11	8	2	3	2	0	1	0	33
14.	G	1953-63	11	9	2	2	2	0	0	0	0
15.	H	1954-63	10	7	2	3	2	0	1	0	33

(continued)

TABLE 8 *(concluded)*

Line	Fore-Cast Set[a]	Period Covered	Number of Years[b]: Covered (NN+TT+TN+NT)	Number of Years[b]: With No TP Observed or Predicted (NN)	Number of Turning Points[b]: Observed (TT+TN)	Predicted (TT+NT)	Correctly Predicted (TT)	Missed (TN)	Falsely Predicted (NT)	Percentage[b] of: Observed Turns Missed[c] ($\bar{E}_{T2}$)	Predicted Turns False[d] ($\bar{E}_{T1}$)
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
						Summary					
16.	8	forecasts of GNP[f]	81	64	15	10	8	7	2	46.7	20.0
17.	7	forecasts of indus. prod.[g]	81	57	17	21	14	3	7	17.6	33.3

[a]The forecasts are described in the text above and Tables 1 and 2. The total period covered by the forecast is used in each case.

[b]For symbols used, see text.

[c]Column 6 as per cent of column 3.

[d]Column 7 as per cent of column 4.

[e]Based on forecasts in constant dollars, as reported, compared with the corresponding actual values.

[f]Figures in columns 1-7 are totals of the corresponding entries on lines 1-8.

[g]Figures in columns 1-7 are totals of the corresponding entries on lines 9-15.

greater here, there are likely to be more false warnings as well. This may help explain the fact that among the annual forecasts of industrial production from the same sources the frequency of false signals is higher than among the GNP forecasts (column 9, lines 16–17).[27]

EVALUATING THE TURNING-POINT RECORD OF RECENT AGGREGATIVE FORECASTS

According to Table 8, forecasters of GNP seem to have failed to predict almost one-half of the turning points that did occur (column 8, line 16). But this hit-and-miss record appears much worse than it is. All these errors refer to the 1957–58 decline, which was very small to begin with in the early annual GNP estimates and was ultimately replaced by a small increase in the current, revised figures (see Table 8, lines 1–2, column 6). Thus, had we used the current instead of the former estimates in our error computations, the record of these forecasts would have been much better on this occasion.

On the other hand, forecasters generally did predict a decline in GNP in 1954 and, until very recently, data bore them out on this score.[28] However, according to the latest revised figures released in July 1965, there was apparently no decrease in the annual GNP series between 1953 and 1954 after all, but rather a minuscule increase (see Table 2, column 2). In this case, then, forecasters were "right" about the sign of change in this series according to all but the most recent data. All this illustrates mainly the highly uncertain nature of directional comparisons when these are applied to a series subject to relatively small short-period changes and revisions large enough to alter the sign of such changes.

Forecasts of industrial production show only three "misses" on seventeen occasions (column 5, line 17). This gives a smaller proportion of errors than that shown for GNP; the results here are more

[27] Of the seven errors of this kind for the industrial production forecasts, five refer to 1961 and one each to 1948 and 1963. In all these cases, declines were predicted but increases actually occurred.

[28] This is shown in Table 2, column 2, where the predicted changes are all negative and so is the actual change according to the first GNP estimates. Later estimates whittled down this decline; those available in May 1965, for example, show a change of −\$2.3 billion instead of −\$7.7 billion, which is the early figure listed in Table 2, line 1.

valid in the sense that they do not depend on the vintage or revisions of the data.[29]

For both GNP and industrial production, all false signals are predictions of peaks (declines) that never materialized; and all missed turns are peaks, i.e., increases were predicted but declines were posted (at least according to the early measurements used here). This situation is also reflected in the distributions of turning-point errors by type of change in the observed figures, as shown in Table 7 above.[30]

These results indicate that the main difficulty in predicting reversals of the economy in the postwar period was with downturns rather than upturns. This is certainly not surprising. In recent times, business contractions have been relatively mild and brief and have been widely expected to be so. They varied much less in duration than expansions did. It was generally assumed that antirecessionary policies would be used to attenuate and cut short any declines in aggregate economic activity. On the other hand, the success of policies designed to steady and lengthen expansions was probably more difficult to gauge.

Furthermore, recent business cycle peaks have occurred early enough in the year to be recognized as such by the end of the year, which is the time when the annual forecasts are made.[31] Thus forecasts for 1954, 1958, and 1961 reflected the widespread assumption that the contractions then in process had already run most of their course.[32]

In short, it is demonstrably quite difficult to appraise annual forecasts

[29] Of these three errors, two refer to the late Korean period (1952) when increases were predicted but declines occurred. One "miss" in the opposite direction is recorded for 1958.

[30] According to Table 7, column 4, all turning-point errors in GNP forecasts since 1952 were associated with declines in GNP. These represent missed peaks. In the industrial production forecasts, six errors were associated with increases and one with a decrease; the former are false signals of peaks, the latter is a missed peak. (Note that Table 7 excludes forecasts made before 1952, which are covered in Table 8.)

[31] According to the National Bureau chronology of business cycle turns, the last three peaks are dated July 1953, July 1957, and May 1960. The corresponding dates of specific peaks in GNP (whether measured in current or constant dollars) are II Q 1953, III Q 1957, and II Q 1960. The downturns in industrial production were considerably earlier on two occasions (February 1957 and January 1960).

[32] This does not imply, however, that the dates of the coming troughs were well specified in these forecasts; it does not even necessarily mean that the dates of the troughs were predicted better than those of the peaks. No such inferences can be made with any assurance from an analysis of annual forecasts; a year is simply too long a unit period for that.

with respect to their turning-point performance, especially for GNP. A year is too long a unit period for such an evaluation, since recessions have been short and mild enough to leave only an uncertain imprint upon the annual data. Nevertheless, an over-all survey suggests a positive answer to the question: Were turns predicted more frequently when observed turns occurred? The evidence lies in the concentration of the recent forecasts of directional change in the two periods during which recessions did actually occur (causing at least the early annual GNP estimates to decline), namely, in 1953–54 and 1957–58.

While this is believed to be a meaningful result in a practical sense, its statistical significance cannot be readily established. The double dichotomy of turning-point forecasts introduced earlier in this discussion invites the application of the familiar chi-square (χ^2) test of independence.[33] When the data in Table 8 are pooled for either GNP or industrial production, high values of χ^2 are obtained, whose acceptance would imply a refutation of the (null) hypothesis that the predicted turning points are randomly distributed.[34] However, such pooling of the data requires independence among the forecast sets, which cannot, in general, be assumed. Forecasters presumably influence each other in various ways and most agree that they do; indeed, the nature of their environment and efforts is such as to make this just about inevitable.

No estimate of the magnitude of these complex effects can be provided with reasonable assurance. Forecasts of the same series, if they are any good, must, of course, be positively correlated. In fact, the correlations among them are typically substantial, as would be expected: after all, these forecasts all draw in part on the same informa-

[33] In our notation, the difference of the cross products in the 2×2 table concerned is $d = NN \times TT - NT \times TN$. Let the marginal totals be $e = NN + NT$, $f = TN + TT$, $g = NN + TN$, and $h = NT + TT$; then $\chi^2 = \frac{d^2n}{efgh}$ (where $n = e + f = g + h$). For small frequencies, the use of a continuous instead of a discontinuous distribution may considerably understate the probability of obtaining the given result by chance. An approximate allowance for this can be made by means of the "Yates' correction," which yields $\chi_y^2 = \frac{(d - n/2)^2n}{efgh}$. See R. A. Fisher, *Statistical Methods for Research Workers*, 12th ed., New York, 1954, Chapter IV.

[34] The values of χ^2 computed from the appropriate entries in lines 16 and 17 of Table 8 are: for GNP, 28.6; for industrial production, 35.7. The corresponding values of χ_y^2 are 24.1 and 32.1. The probabilities of obtaining such results by chance are very small indeed (less than .001).

tion and knowledge about the past behavior of the given series and related variables. This holds even if the forecasters did not influence each other directly at all. Therefore, the effects of any such influences, whatever their cause (reputation of some leading experts, pressures to conform, contagious expectations, etc.), would probably be to strengthen the correlations that already exist. Lacking detailed insight into each forecaster's procedures, it seems impossible to separate the effects of the common data from those direct influences that are opposed to the hypothesis of forecasters' mutual independence.[35]

Comparison of forecasts does reveal considerable diversity, however, suggesting a large role for independent analysis and individual judgment, despite the undoubtedly important common elements (see, e.g., Chart 1). Pure imitation must be rare among self-respecting forecasters, and the observed similarities are usually partial and temporary. Pooling the forecasts from different sources, therefore, will presumably result in a certain effective increase in the number of independent observations. But this increase is certainly less than that implied in the simple sums used in the over-all χ^2 test; hence the latter understates, by an unknown quantity, the probability of obtaining the observed results by chance. Consequently, the results of such tests (see footnote 34) may at best be suggestive, but cannot be interpreted in any strict sense.

If the periods covered were longer, so that the numbers of observations were sufficiently large, the tests of whether predicted turns are randomly distributed or associated with actual turns could be usefully applied to data for each separate forecast set. As it is, however, the turning-point frequencies for the individual sets included in Table 8 are too small to permit reliable tests of this sort.

[35] One might speculate that common successes are attributable to the use of similar, valid methods in handling the same data, while common errors are due to direct influences such as the spread of plausible expectations that proved wrong. But this need not be so. It is true, of course, that forecasts may share errors as well as successes; their concordance does not always signify a higher correlation with the actual outcomes. But it is possible for an analysis recognized as valid and used independently by many forecasters to lead occasionally to widespread errors; and, conversely, for an influential individual's view of the future, which was adopted by many, to prove right.