NBER WORKING PAPER SERIES

THE ACCURACY OF INDIVIDUAL AND GROUP FORECASTS FROM BUSINESS OUTLOOK SURVEYS

Victor Zarnowitz

Working Paper No. 1053

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138

December 1982

Financial support from the National Science Foundation and aid of the Graduate School of Business of the University of Chicago and NBER are gratefully acknowledged. I am greatly indebted to Louis Lambros for helpful suggestions and efficient research assistance. This paper contains revised parts of the NBER Working Paper No. 845. The research reported here is part of the NBER's research program in Economic Fluctuations. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper #1053
December 1982

The Accuracy of Individual and Group Forecasts from Business Outlook Surveys

Abstract

This paper reports on a comprehensive study of the distributions of summary measures of error for a large collection of quarterly multiperiod predictions of six variables representing inflation, real growth, unemployment, and percentage changes in nominal GNP and two of its more volatile components. The data come from surveys conducted since 1968 by the National Bureau of Economic Research and the American Statistical Association and cover more than 70 individuals professionally engaged in forecasting the course of the U. S. economy (mostly economists, analysts, and executives from the world of corporate business and finance). There is considerable differentiation among these forecasts, across the individuals, variables, and predictive horizons covered. Combining corresponding predictions from different sources can result in significant gains; thus the group mean forecasts are on the average over time more accurate than most of the corresponding sets of individual forecasts. But there is also a moderate degree of consistency in the relative performance of a sufficient number of the survey members, as evidenced in positive rank correlations among ratios of the individual to group root mean square errors.

Victor Zarnowitz
University of Chicago
Graduate School of Business
1101 East 58th Street
Chicago, Illinois 60637

(312) 962-7130

I. Introduction

Empirical studies of forecasts and expectations based on survey data have generally concentrated on the performance of time series of averages of the participants' responses. As a rule, these represent means or medians for groups whose size and composition vary over time. This raised the possibility of serious aggregation errors due to the neglect of the cross-section and distributional aspects of the data: differences among the individuals and subgroups; sampling variation; consistency and representativeness of the employed averages. That such matters can be important is not in doubt, but they seem to have attracted relatively little attention in the literature.[1]

This paper examines the accuracy of a large number of individual forecast series and of the corresponding average forecast series from a quarterly survey conducted by the author for the National Bureau of Economic Research in collaboration with the American Statistical Association. The survey questionnaire is mailed by the ASA in the middle month of each quarter to a list of persons who are professionally engaged in forecasting the course of the economy, and regular reports on the results are released in the third month.[2] The respondents are economists, mainly from corporate business and finance but also from independent consulting firms, governmemt agencies, academic and research

---

[1]An early study which dealt with certain characteristics of the relation between aggregate and individual forecasts is Zarnowitz, 1967, pp. 123-126. A recent analysis of disaggregated data from surveys of inflation forecasts is Figlewski and Wachtel, 1981.

[2]The reports, prepared by the NBER, are now published in the NBER Reporter, and in AmStat News. They discuss mainly the median predictions of current interest. For some of the broader historical evaluations, see Moore, 1969, 1977; Zarnowitz, 1972, 1979; Fair, 1974; Christ, 1975; McNees, 1975, 1976; Su and Su, 1975.

organizations. This study covers 79 individuals (persons or firms) who participated in at least 12 of the 42 surveys in the period from 1968:4 through 1979:1.[3]

The forecasts relate to rates of change in four variables: gross national product in current and constant dollars (labeled GNP and RGNP); the GNP implicit price deflator (IPD); and consumer expenditures for durable goods (CEDG). The errors of percentage change forecasts are

$$(1) \quad e_{t+j} = \begin{cases} \left(\dfrac{P_t - A^*_{t-1}}{A^*_{t-1}} - \dfrac{A_t - A_{t-1}}{A_{t-1}}\right) 100 , & \text{if } j = 0 \\[2ex] \left(\dfrac{P_{t+j} - P_{t+j-1}}{P_{t+j-1}} - \dfrac{A_{t+j} - A_{t+j-1}}{A_{t+j-1}}\right) 100 , & \text{if } j = 1, \ldots, 4 . \end{cases}$$

Here P is the predicted level and A is the actual level according to the last national income and product accounts data released prior to the major benchmark revisions of January 1976 and December 1980. A^*_{t-1} is a preliminary estimate which is the most recent "actual" value available at the time of the forecast (since A_t is unknown, P_t is a true prediction with a horizon of about one quarter). The subscripts refer to the survey quarter t, which is the date when the forecast was made, and to the target quarter, $t + j$, which is the date to which the forecast refers (since all this applies to any of the forecasters and to any of the variables covered, other subscripts are omitted).

It will be noted that (1) contains differences between the successive levels predicted in a multiperiod forecast made at time t, namely $P_t - A^*_{t-1}$ for the current quarter ($j = 0$) and $P_{t+j} - P_{t+j-1}$ for any of the next four future quarters ($j = 1, \ldots, 4$). Accordingly, these are errors of the implicit

[3]For further discussion and analysis of the ASA-NBER forecast data, and references to the literature, see Zarnowitz, 1983.

marginal or "intra-forecast" change predictions whose targets are successive quarterly intervals (0-1, 1-2, ...), which do not overlap.[4]

For two variables, change in business inventories (CBI) and the unemployment rate (UR), the forecast errors are defined as

$$E_{t+j} = P_{t+j} - A_{t+j}\ , \quad j = 0, 1, \ldots, 4\ , \tag{2}$$

that is, as differences, predicted level minus actual level. These series, unlike the others which have strong upward trends, can be treated as stationary. Here it is the levels that are of primary interest, not the rates of change as in the cases of RGNP (real growth) and IPD (inflation).

The questions addressed are the following: How accurate are the individual forecasts relative to the corresponding group averages? How representative are the latter of the former? What are the distributions across the individuals of the summary measures of error for the period covered? How do the results compare across the different variables and predictive horizons? The paper is a progress report on a comprehensive study of a large and diversified collection of U. S. macroeconomic predictions; other aspects of forecasting behavior and performance will be taken up in other papers.[5]

---

[4] In contrast, forecasts of average changes over increasing spans (0-1, 1-2, ...) have overlapping target periods, and they are therefore necessarily intercorrelated. On the definitions, measures, and merits of level and change errors, see Zarnowitz, 1967, pp. 32-35, now 1979, p. 6, and McNees, 1973, pp. 7-10.

[5] For a report on tests of bias or "rationality," see Zarnowitz, 1983. Sequels will deal with the variations over time, cross-sectional (survey-by-survey) results, disaggregation by method, and probabilistic predictions.

II. Measures of Relative Accuracy and Consistency

The root mean square error (RSME) of the i^{th} individual's set of predictions can be written as

$$M_i = \left(\frac{1}{n_i} \sum_{t \in N_i} \varepsilon_{it}^2\right)^{1/2} \tag{3}$$

for any variable and forecast horizon.[6] Here $\{N_i\}$ is the set of the target periods of the i^{th} forecasts, while n_i is the number of predictions in that set. The numbers and dates of the surveys covered differ across the individuals, and the error series ε_{it} in (3) have gaps at times when any of the forecasters missed any of the surveys, which happened frequently.[7]

Next we construct series of group means predictions that match the series for each individual precisely in terms of the variable, horizon, and periods covered. Thus for each series of predictions by a particular forecaster (denoted by the subscript "i") there is now a corresponding series of group averages ("q_i") of predictions by all those forecasters in our sample who responded to the same surveys. In our simplified notation, the RMSE for the group mean series is

$$M_{qi} = \left(\frac{1}{n_i} \sum_{t \in N_i} \varepsilon_{qt}^2\right)^{1/2} \tag{4}$$

---

[6]For level forecasts (UR and CBI) $\varepsilon_{it} = E_{it}$; for percentage change forecasts (the other variables) $\varepsilon_{it} = e_{it}$. Again there is no need here to complicate the formula by adding subscripts for the variable and target period.

[7]Recall that, to be included, a forecaster must have participated in at least 12 surveys, but the surveys need not be consecutive. The mean number of surveys covered is 23, with a standard deviation of 8; the minimum is 12, the maximum 37 (out of a total of 42).

Ratios of root mean square errors, M_i/M_{qi} provide convenient measures of relative accuracy of individual forecasts. They are comparable in a way in which absolute errors for sets of predictions that differ in target dates are not. The group averages represent an important benchmark of forecasting performance that is reasonably accessible and, as shown below, comparatively efficient. Related data on the median forecasts from the ASA-NBER surveys are summarized after each survey and published regularly, after having been first communicated to the survey members: they reflect the views of many respected professional forecasters and are among the best known and most used predictions for the U. S. economy.

The M_i/M_{qi} Ratios: Individuals vs. Group Averages

Inspection of graphs for 30 distributions of ratios of root mean square errors, M_i/M_{qi} (one for each of the six variables and five target quarters) shows that every one of them is skewed to the right. This is illustrated in Chart 1, which includes the graphs for the shortest horizon ($j = 0$) and the same-quarter-year-ahead target ($j = 3$), those periods being labeled Q0 and Q3, respectively. It is clear that only minorities of the individuals had ratios of less than 1, that is, outperformed the group averages over time.

Summing up the evidence from all such graphs (for Q0, ..., Q4), the best (lowest) ratios fall between 0.7 and 0.9, the worst (highest) between 1.4 and 2.2. The means of the ratios (marked $\overline{M}$) are all located to the right of the unity (broken vertical) lines. The histograms tend to get tighter and also, often, less skewed for the more distant quarters.[8]

---

[8] See Zarnowitz, 1982, pp. 16-18, for a chart showing all the graphs discussed in the text above.

CHART 1

SEVENTY-NINE INDIVIDUAL FORECASTS OF MULTIPERIOD CHANGES IN SIX AGGREGATE VARIABLES, COMPARISONS WITH GROUP MEAN FORECASTS, 1968-1979

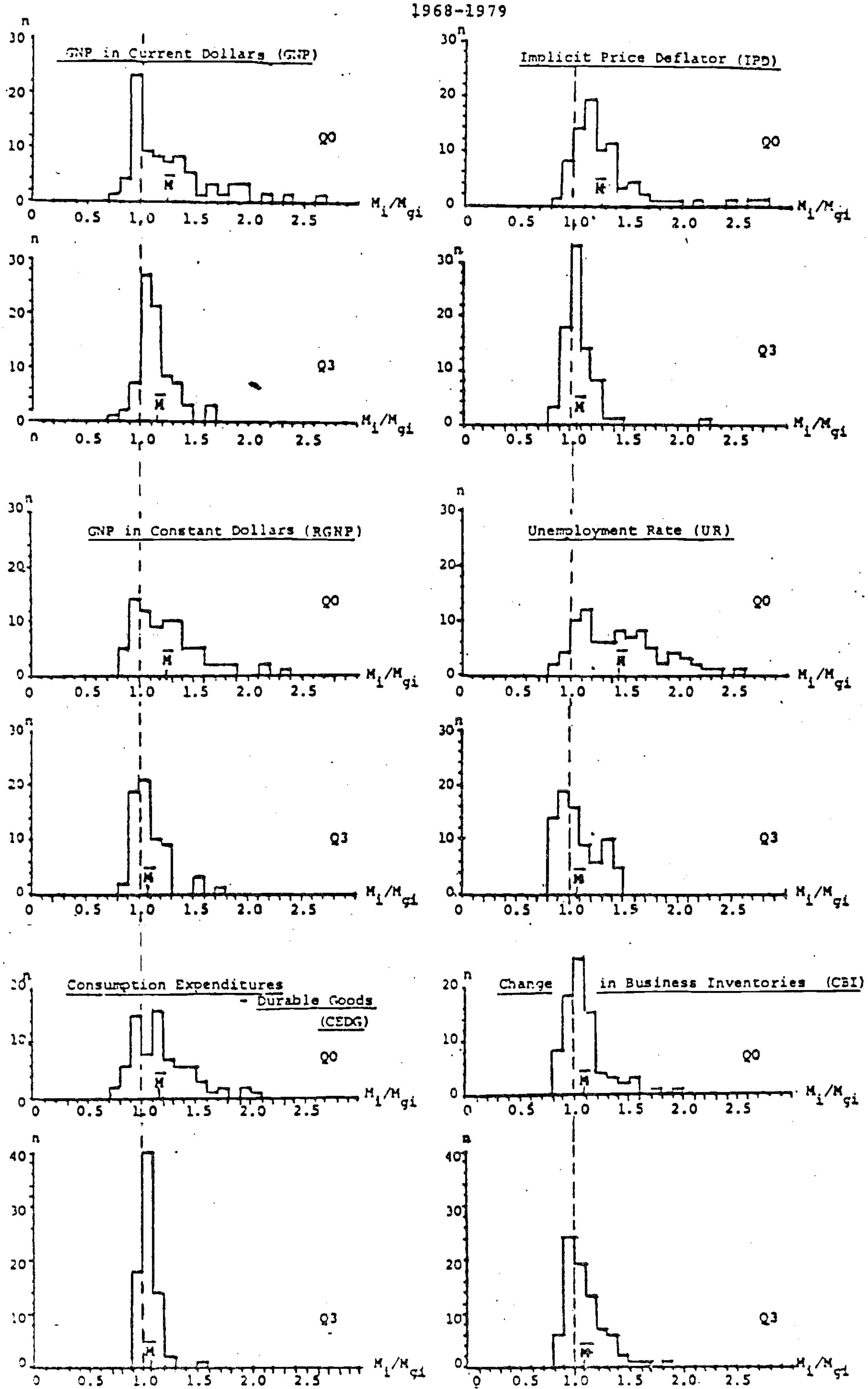

Table 1 shows that the mean ratios are remarkably close: when rounded, all but nine of the 30 statistics are 1.1. The higher mean ratios, ranging from 1.2 to 1.4, refer to the shortest predictions, for Q0 and, less so, for Q1. The standard deviations of the M_i/M_{gi} ratios tend to decreaase strongly with the distance to the target quarter, from Q0 to Q3.[9] An exception is CBI, where the horizon of the expectations apparently does not matter much (all the means are approximately 1.1 and the decline in the dispersion of the ratios is very small).

The proportions of the better-than-average forecasters ($M_i/M_{gi} < 1$) vary strongly with the target quarter for some variables, much less so for others. Thus for UR the range is 8 to 42 percent, for CBI it is only 29 to 38 percent. Averaged across Q0-Q4, the figures fall between 20 percent for GNP and 33 percent for CBI (see the last section of Table 1).

It is known from past studies (and shown again below) that the average accuracy of forecasts varies considerably across the individuals, variables, and target periods. Highly volatile series such as CEDG and CBI are much more difficult to predict than relatively smooth, trend-dominated series such as GNP. In general, the uncertainty and difficulty (hence errors) of prediction tend to increase for the more distant future. The remarkable degree of standardization in the M_i/M_{gi} ratios stands in sharp contrast to the diversity of the average accuracy measures for the individuals, M_i.

---

[9]The series for Q0, Q1, Q2, and Q3 start in 1968:4, 1969:1, 1969:2, and 1969:3, respectively, and extend through 1979:1. The series for Q4 start in 1969:4 and end in 1979:1 but miss the first three quarters in 1970, 1971:1, and 1975:3 (because a few surveys did not ask for the Q4 predictions). For these reasons, the number of the surveys covered is 42 for Q0, 41 for Q1, 40 for Q2, 39 for Q3, and 33 for Q4. Our comparisons are somewhat impaired by these disparities; in particular, the relatively large figures for Q4 compared to those for Q3 probably reflect the drop in survey coverage.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF THE M_i/M_{gi} RATIOS, BY VARIABLE AND TARGET QUARTER, 1968-1979

	GNP (1)	IPD (2)	RGNP (3)	UR (4)	CEDG (5)	CBI (6)
				MEANS		
Q0	1.25	1.29	1.25	1.43	1.18	1.10
Q1	1.16	1.16	1.14	1.19	1.10	1.11
Q2	1.13	1.11	1.11	1.11	1.07	1.09
Q3	1.14	1.08	1.09	1.08	1.06	1.10
Q4	1.17	1.10	1.12	1.08	1.08	1.10
				STANDARD DEVIATIONS		
Q0	.38	.36	.31	.36	.28	.20
Q1	.21	.23	.18	.24	.18	.21
Q2	.18	.16	.15	.19	.11	.19
Q3	.17	.18	.15	.18	.09	.18
Q4	.24	.20	.19	.19	.10	.19
				PERCENTAGE OF CASES WHERE $M_i/M_{gi} < 1$		
Average	20	26	22	29	24	33
Range	13-35	11-37	18-27	8-42	19-31	29-38

Note: Based on quarterly ASA-NBER business outlook surveys 1968:4-1979:1. On coverage and symbols used, see text and Chart 1.

The advantage of the group means M_{gi} is the greatest for the nearest targets and it becomes less and less important as the predictions reach out further into the future. One may speculate that the individual forecasts for Q0 and Q1 contain more independent information than those for Q2-Q4, hence the gains from averaging are larger for the former than for the latter.[10] The abilities to predict CBI are particularly limited, even for the nearest quarters, so here the means and dispersion of the ratios M_i/M_{gi} depend little on the distance to the target quarter (j = 0, 1, . . ., 4).

Earlier data, on predictions by members of a large group of business economists organized into the New York Forecasters' Club, produce similar results. The distributions of M_i/M_{gi} ratios for six-month and twelve-month forecasts of industrial production in 1947-63 show strong positive skewness, with most of the values falling between 1.0 and 1.3, the classes below 0.8 almost empty, and the average values all concentrated in the narrow range, 1.1-1.2[11]

These findings indicate that it is difficult for most individuals to predict consistently better than the group. Contemporaneous expectations for a given target may be distributed more or less symmetrically about their mean, but over time the individuals' positions within these distributions are likely

---

[10]The large means and standard deviations of the ratios for Q0 may be associated with the disparities in the quality of the current data available to different individuals. Although the survey questionnaire provides the most recent information on the values of the series to be predicted, some respondents choose to use different jump-off levels which may be more or less accurate. It is not quite clear why the figures for the shortest predictions of UR should be particularly high, as Table 1, column 4, shows them to be, but it is suggestive that this is the only variable covered for which monthly data are available. Some individuals are likely to lag behind the majority in absorbing these monthly data (and related weekly information on unemployment claims).

[11]See Zarnowitz, 1982, p. 22, for a chart showing these results in detail. GNP forecasts made by members of the same group in the period 1956-63 (Zarnowitz, 1967, pp. 123-126), tell much the same story.

to fluctuate. For most people, most of the time, the predictive record is spotty, with but transitory spells of relatively high accuracy. A series of group averages has the advantage that it is helped by the cancellation of individual errors of opposite sign.

Rank Tests of Predictive Consistency

Success in one class of predictions (say, for GNP in Q1) may or may not coincide with success in another class (say, GNP in Q4, or for IPD). If the degree of coincidence were very low (e.g., if very few people managed to "beat" the group mean in more than one class), then the success, being rather isolated, might be attributable more to chance than to better techniques or skills.

The NBER-ASA survey participants have been ranked according to the M_i/M_{gi} ratios for each of the variables and target quarters covered. The correlations among the resulting ranks could be either close to zero (indicating very little consistency in the relative performance of the forecasters across different variables or predictive spans) or significantly negative (those who succeed in one category tend to fail in another) or significantly positive (those who succeed in one category also tend to succeed in others).

The rank correlations are presented in Table 2, both across the variables for each target quarter (part I) and across target quarters for each variable (part II). All the correlations are positive and in general they appear to be significantly so (see note in the table). Thus there is some degree of consistency in the predictive performance of the individuals as revealed by their M_i/M_{gi} ranks.

People who predict relatively well the rates of change in nominal GNP also tend to do so for the rates of change in real GNP: the average rank correlation coefficient $\overline{\rho}$ is 0.74 is this case. For variables that are not so

TABLE 2

RANK CORRELATIONS AMONG PARTICIPANTS IN ASA-NBER SURVEYS ACCORDING TO RATIOS OF INDIVIDUAL TO GROUP ROOT MEAN SQUARE ERRORS, MULTIPERIOD PREDICTIONS FOR SIX AGGREGATE VARIABLES, 1968-1979

I. Across Variables, for Each Target Quarter[a]

Q0	GNP	IPD	RGNP	UR	CEDG	CBI
GNP	1.00					
IPD	.57	1.00				
RGNP	.83	.65	1.00			
UR	.42	.43	.39	1.00		
CEDG	.69	.56	.69	.50	1.00	
CBI	.40	.41	.36	.21	.42	1.00

Q1	GNP	IPD	RGNP	UR	CEDG	CBI
GNP	1.00					
IPD	.23	1.00				
RGNP	.77	.48	1.00			
UR	.20	.30	.40	1.00		
CEDG	.43	.20	.50	.30	1.00	
CBI	.39	.31	.38	.22	.38	1.00

Q2	GNP	IPD	RGNP	UR	CEDG	CBI
GNP	1.00					
IPD	.35	1.00				
RGNP	.66	.48	1.00			
UR	.23	.48	.27	1.00		
CEDG	.27	.15	.21	.12	1.00	
CBI	.41	.43	.32	.36	.31	1.00

Q3	GNP	IPD	RGNP	UR	CEDG	CBI
GNP	1.00					
IPD	.27	1.00				
RGNP	.69	.33	1.00			
UR	.41	.44	.49	1.00		
CEDG	.21	.05	.14	.19	1.00	
CBI	.28	.42	.15	.37	.03	1.00

Q4	GNP	IPD	RGNP	UR	CEDG	CBI
GNP	1.00					
IPD	.44	1.00				
RGNP	.76	.51	1.00			
UR	.39	.49	.31	1.00		
CEDG	.18	.17	.17	.06	1.00	
CBI	.49	.36	.38	.19	.21	1.00

Average, Q0-Q4	GNP	IPD	RGNP	UR	CEDG	CBI
GNP	1.00					
IPD	.37	1.00				
RGNP	.74	.49	1.00			
UR	.33	.45	.35	1.00		
CEDG	.36	.23	.34	.23	1.00	
CBI	.32	.39	.32	.27	.27	1.00

[a]These measures refer to 75 individuals who participated in at least 12 quarterly ASA-NBER business outlook surveys 1968:4-1979:1 and predicted all six variables covered. The symbols for the variables are identified in part II of the table. The symbols Q0-Q4 refer to the current and the four successive future quarters. The rank correlation coefficients shown are Spearman's $\rho = 1 - [6\Sigma d^2/(n^3 - n)]$ where d is the rank difference and n is the number in each ranking.

TABLE 2
(concluded)

II. Across Target Quarters, for Each Variable[b]

GNP in Current Dollars (GNP)

	Q0	Q1	Q2	Q3	Q4
Q0	1.00				
Q1	.51	1.00			
Q2	.18	.19	1.00		
Q3	.14	.18	.40	1.00	
Q4	.50	.32	.40	.47	1.00

Implicit Price Deflator (IPD)

	Q0	Q1	Q2	Q3	Q4
Q0	1.00				
Q1	.55	1.00			
Q2	.45	.68	1.00		
Q3	.41	.54	.60	1.00	
Q4	.39	.52	.51	.62	1.00

GNP in Constant Dollars (RGNP)

	Q0	Q1	Q2	Q3	Q4
Q0	1.00				
Q1	.57	1.00			
Q2	.33	.47	1.00		
Q3	.05	.19	.48	1.00	
Q4	.38	.19	.33	.33	1.00

Unemployment Rate (UR)

	Q0	Q1	Q2	Q3	Q4
Q0	1.00				
Q1	.64	1.00			
Q2	.38	.78	1.00		
Q3	.32	.62	.85	1.00	
Q4	.27	.53	.75	.92	1.00

Consumer Expenditures-Durable Goods (CEDG)

	Q0	Q1	Q2	Q3	Q4
Q0	1.00				
Q1	.66	1.00			
Q2	.53	.43	1.00		
Q3	.12	.07	.14	1.00	
Q4	.14	.00	.04	.08	1.00

Change in Business Inventories (CBI)

	Q0	Q1	Q2	Q3	Q4
Q0	1.00				
Q1	.70	1.00			
Q2	.63	.76	1.00		
Q3	.51	.56	.81	1.00	
Q4	.57	.62	.63	.76	1.00

---

[b] These measures refer to the sample covered in Chart 2: 79 individuals for each of the variables except CEDG (80). The rank correlation coefficients are Spearman's ρ.

NOTE: For rankings without ties, the variance of ρ equal's $\frac{n}{n-1}$ (Kendall 1948, p. 46). For $n = 75$, therefore, the standard error $S_\rho = 1/\sqrt{0.74} = 0.1162$ (for $n = 79$, $S_\rho = 0.1125$; for $n = 80$, $S_\rho = 0.1132$). Hence, all entries $\rho \geq 0.23$ in the table are significant at the 5% level, and all $\rho \geq 0.20$ at the 10% level.

closely related, the correlations are much lower (e.g., $\bar{\rho}$ = 0.23 for CEDG and IPD, and also for CEDG and UR). However, only 15 of the 75 coefficients ($\rho \neq 1$) in part I of the table are less than 0.2. The overall mean of the ρ statistics is 0.36.

For any of the variables, people who rank high (low) in predicting one quarter also tend to rank high (low) in predicting the next quarter. The ρ's for Q0-Q1 average 0.61, those for Q1-Q2, Q2-Q3, and Q3-Q4 average 0.52-0.55 (see part II of Table 2). For non-adjoining target periods, the rank correlations are lower, $\bar{\rho}$ being 0.40 where the distance is two quarters (Q0-Q2, Q1-Q3, and Q2-Q4) and 0.31 where it is three quarters (Q0-Q3 and Q1-Q4). The further apart the target periods, the less correlated are the values to be predicted, and the above results suggest that the ranking consistency declines correspondingly. But the reductions in the rank correlations vary considerably in size and regularity, being most pronounced for CEDG, least for GNP. When averaged over the quarters Q1-Q4, the $\bar{\rho}$ coefficients are relatively low for CEDG, GNP, and RGNP (.27-.33) and high for IPD, UR, and CBI (.55-66).

III. Distributions of Summary Measures of Error

It is instructive to examine the distributions of the statistics that sum up the records of the individual forecasters. The discrepancies in time coverage reduce the comparability of absolute accuracy measures across the respondents to the surveys. However, in the ASA-NBER data there appears to be no significant bias due to missed observations. No pattern has been found to suggest that the participants covered selected the times of their responses in any systematic manner; rather it is random factors (absences, work pressure, negligence) that account for the allocation of the missed surveys among the individuals. Interest in the overall picture provided by the summary measures of each forecaster's performance is also enhanced by the fact that the number of surveys (42) is relatively large and the coverage of each is adequate (on the average, 43 participants with a standard deviation of 9).[12]

The distributions of the summary measures of error for the individual forecasts are further compared with the corresponding measures for the overall group forecasts. The latter refer to the series of mean predictions, of which there are thirty, one for each of the targets covered (6 variables × 5 horizons). These averages comprise all forecasters who predicted the given target at any time during the period under study, so that the series are continuous, each including predictions from all surveys covered. Thus the RMSE for any of these group mean ("g") series is simply

$$(5) \qquad M_g = \left(\frac{1}{n} \sum_t \varepsilon_{gt}^2\right)^{1/2}, \qquad t = 1, 2, \ldots, n,$$

where n is the total number of consecutive surveys ($42 - j$ for Q0-Q3, 33 for Q4, see note 9).

Overall Accuracy

For each of the six variables, the means of the individual RMSE's taken across the target quarters Q0, ...,Q4 exceed the corresponding RMSE's for the overall group mean forecasts. The ratios of the summary statistics of error (entries in column 2 of Table 3 divided by those in column 5) vary from 1.04 to 1.16 and average 1.11.

The performance of the series of group mean forecasts is also superior to the average performance of the series of individual forecasts in terms of correlations with the actual values. The averages of the $\bar{r}^2$ coefficients

---

[12]See Zarnowitz, 1983, for more numerical detail on the forecast samples from the ASA-NBER surveys.

Table 3

SELECTED OVERALL ACCURACY STATISTICS FOR INDIVIDUAL AND GROUP MEAN FORECASTS, SIX VARIABLES, 1968-1979

Variable[a]	Individual Forecasts[b]			Group Mean Forecasts[c]			Actual Values[d]		
	ME	RSME	$\bar{r}^2$	ME	RSME	$\bar{r}^2$	Mean	SD	RMSV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
GNP	-.11	1.00	.22	-.13	.88	.29	2.18	1.09	2.93
IPD	-.39	.78	.21	-.38	.67	.27	1.50	.67	1.64
RGNP	.28	1.21	.28	.25	1.05	.35	.68	1.24	1.41
UR	-.14	.67	.66	-.10	.62	.69	5.78	1.68	6.02
CEDG	-.28	4.04	.14	-.36	3.68	.18	2.25	3.97	4.57
CBI	-1.72	10.00	.27	-1.77	9.57	.44	9.19	10.87	14.23

[a]On the symbols used, see note d below.

[b]These measures refer to the sample covered in Chart 1 (75 individuals forecast CEDG, 79 each of the other variables). They are means of the corresponding statistics for the five target quarters, Q0, ..., Q4. ME = mean error; RMSE = root mean square error; $\bar{r}^2$ = squared coefficient of correlation, corrected for the degrees of freedom.

[c]These measures refer to the overall group mean forecasts M_g (see eq. 5 and text) and are means of the corresponding statistics for the target quarters Q0, ..., Q4). See note b above for the explanation of the symbols.

[d]For the definition of actual values, see text. For the nominal and real gross national product (GNP and RGNP), the GNP implicit price deflator (IPD), and nominal consumer expenditures on durable goods (CEDG), the measures refer to percentage changes; for the unemployment rate (UR) and the change in business inventories (CBI), they refer to levels. SD = standard deviation (corresponding to the means in column 7); RMSV = root mean square value computed as $(\text{mean})^2 + (\text{SD})^2$.

for the individuals vary between .14 and .28, except for the unemployment rate, a relatively smooth level series, where the $\bar{r}^2$ is .66 (column 3). They are 60 to 80 percent lower than their counterparts for the overall group means (column 6), except again for UR, where the margin in favor of the aggregate is much smaller.

The mean errors have negative signs for all the variables, with the important exception of RGNP. This reflects the familiar tendency toward underestimation of changes in most forecasts. The average overestimation of real growth observed in our data is largely explained by the fact that, after a decade of relative stability and an extraordinarily long business expansion, the 1970s gave rise to a novel phenomenon commonly called stagflation and an unexpectedly serious recession. As would be expected, since the individual predictions are randomly distributed over the same period as that covered by the overall group mean series, the two sets of forecasts have much the same mean errors (cf. columns 1 and 4).

As a rule, it is some simple average rather than the underlying individual forecasts from economic outlook that are regularly published and used, and it is certainly worth knowing that the predictive value of the former tends to be measurably greater than that of the latter; but how accurate have the mean predictions been, considering the accessible data and techniques? One approach to answering this broad question would be through comparisons with benchmark predictions from time-series models appropriately selected to fit the characteristics of the variables in question and estimated with data available at the time the ex ante forecasts to be assessed were actually made. This task is beyond the scope of the present paper. To gain some insight into the order of the magnitudes involved, however, it is useful to compare the average forecast errors with the average values of the outcomes for each of the target series, and some summary statistics are provided for this purpose in the last section of Table 3.

Plainly, the absolute values of the mean errors are at least smaller than the mean actual values in every case, and they are indeed for most of the variables quite small in these terms (cf. columns 1 and 4 with column 7). More telling, the RMSE's are less than the corresponding root mean square values of the target series, again in most cases by large margins (cf. columns 2 and 5 with column 9). The RSME's for the group mean forecasts are also generally less than the standard deviations of the actual values (columns 5 and 8). The predictions of UR and GNP rank as the first and second best in all of these comparisons; IPD and RGNP rank lowest when the ME figures are used, RGNP and CEDG when the RMSE figures are.

Characteristics of the Distributions

The medians of the RMSE's for the individual forecasts are with few exceptions lower than the means, but by relatively small margins (see Table 4, columns 1 and 4). This indicates a weak tendency for these distributions to be skewed to the right, that is, toward the large RMSE's.

In virtually all instances, the averages of the individual RMSE's exceed the RMSE's for the corresponding group mean forecasts (compare the entries in columns 1 and 4 with their counterparts in column 6). The measures for the group mean tend to be closer to the lower quartile than to the median of the distribution of the individual RMSE's (cf. columns 3, 4, and 6). This is roughly consistent with the earlier finding, based on more strictly comparable measures, that the overall proportion of cases in which $M_i < M_{qi}$ is about 26 percent (Table 1).

The more distant the target quarter, the larger tend to be the prediction errors, as demonstrated by the increases from Q0 through Q4 of the entries in

TABLE 4

ROOT MEAN SQUARE ERRORS AND CORRELATIONS BETWEEN PREDICTED AND ACTUAL VALUES, SELECTED DISTRIBUTIONAL STATISTICS BY VARIABLE AND TARGET QUARTER, INDIVIDUAL AND GROUP MEAN FORECASTS, 1968-1979

Quarter Predicted	Root Mean Square Errors (RMSE)						Squared Correlations ($\bar{r}^2$)		
	Individual Forecasts					Group Mean	Individual Forecasts		Group Mean
	M (1)	SD (2)	LQ (3)	Me (4)	UQ (5)	M_g (6)	M (7)	SD (8) (9)	M_g
				GNP in Current Dollars (GNP)					
Q0	.77	.27	.58	.68	.92	.66	.49	.22	.63
Q1	.95	.26	.78	.92	1.05	.86	.24	.16	.31
Q2	1.06	.19	.96	1.07	1.15	.96	.13	.12	.18
Q3	1.10	.26	.92	1.10	1.23	.98	.12	.13	.16
Q4	1.12	.28	.94	1.08	1.22	.94	.12	.15	.18
				Implicit Price Deflator					
Q0	.55	.16	.45	.49	.60	.42	.45	.20	.64
Q1	.69	.16	.58	.66	.77	.59	.28	.18	.35
Q2	.79	.16	.69	.78	.87	.70	.14	.12	.17
Q3	.88	.19	.78	.86	.95	.77	.10	.10	.12
Q4	.98	.21	.86	.94	1.09	.88	.10	.10	.08
				GNP in Constant Dollars (RGNP)					
Q0	.85	.28	.67	.78	.96	.70	.60	.18	.75
Q1	1.09	.28	.91	1.03	1.26	.95	.38	.17	.48
Q2	1.24	.22	1.13	1.25	1.37	1.12	.18	.13	.25
Q3	1.39	.25	1.22	1.36	1.53	1.23	.10	.11	.10
Q4	1.46	.31	1.25	1.39	1.69	1.23	.12	.15	.16
				Unemployment Rate					
Q0	.22	.06	.17	.21	.26	.16	.97	.02	.99
Q1	.46	.11	.38	.44	.51	.41	.86	.06	.91
Q2	.71	.17	.60	.67	.81	.65	.68	.12	.75
Q3	.94	.23	.78	.91	1.09	.88	.48	.17	.53
Q4	1.04	.24	.88	1.00	1.19	.98	.32	.19	.27
				Consumer Expenditures--Durable Goods (CEDG)					
Q0	3.37	.96	2.64	3.10	3.85	2.87	.40	.25	.63
Q1	4.16	.78	3.79	4.13	4.60	3.77	.11	.15	.13
Q2	4.24	.71	3.82	4.22	4.68	4.04	.05	.07	.01
Q3	4.44	.65	4.12	4.41	4.78	4.09	.05	.07	.01
Q4	3.98	.69	3.47	3.99	4.47	3.64	.08	.08	.13
				Change in Business Inventories (CBI)					
Q0	8.21	2.65	6.70	8.10	9.61	8.07	.36	.21	.55
Q1	9.17	3.06	7.06	8.89	11.86	9.11	.31	.22	.51
Q2	10.42	3.36	8.22	10.08	12.87	9.79	.25	.19	.41
Q3	10.99	3.32	9.16	11.12	13.10	10.08	.20	.18	.40
Q4	11.22	3.00	9.13	11.14	13.12	10.80	.21	.17	.35

NOTE: The measures refer to those individuals who participated in at least 12 of the quarterly ASA-NBER business outlook surveys in 1968:4 -1979:1 (75 for CEDG, 79 for each of the other variables). Q0 denotes the current (survey) quarter, Q1-Q4 the following four quarters (for details on coverage by target quarter, see note 9). M denotes mean; SD, standard deviation, LQ, lower quartile; UQ, upper quartile, and M_g, overall group mean (see eq. 5 and text). The $\bar{r}^2$ are corrected for the degrees of freedom.

columns 1 and 3-6 of Table 4. However, the increases taper off: the forecasters on the average predict Q0 substantially better than Q1, and Q1 still noticeably better than Q2, but their ability to anticipate Q3, is not much less limited than their ability to anticipate Q2, and the same applies even more to Q4 vs. Q3. In short, these measures suggest that the RMSE's tend to approach asymptotically a high plateau at the more distant target quarters.

Note that these results apply to the marginal prediction errors for each successive quarter (in a shorthand notation used earlier, to changes 0-1, 1-2, ...). To the extent that such errors are positively correlated, their cumulation can produce much greater increases in the average prediction errors for changes over increasing, overlapping spans (0-1, 0-2, ...).[12]

The absolute dispersion measures (standard deviations in column 2 and interquartile ranges implied by columns 3 and 5) increase from Q0 to Q4 for UR, decrease for CEDG, and behave rather irregularly for other variables such as GNP and RGNP. In contrast, relative dispersion measures, viz., the coefficients of variation SD/M (ratios of entries in column 2 to those in column 1) show strong tendencies to decrease for the more distant target quarters. They are also on the average similar for most of the variables (ranging from .23 to .26, except for CEDG and CBI, where they are .19 and .31, respectively).

---

[12]The buildup of average prediction errors with increasing spans is a general phenomenon to be expected and is well documented in forecast evaluations. However, some evidence for earlier periods has shown marginal errors varying narrowly and irregularly over the range of several quarters ahead, without any systematic upward drift (Zarnowitz 1967, pp. 64-72, and 1979, pp. 18-19; McNees 1973, pp. 24-25). The present results may differ because of the nature of the period covered (and Zarnowitz 1979 provides some support for this hypothesis), but they also inspire more confidence than those of other studies, being based on much larger samples of better controled data.

The group mean forecasts have tracked the actual changes better than the average individual forecasts: the correlation measures in column 9 of Table 4 are, with but a few exceptions, higher than those in column 7. The listed $\bar{r}^2$ coefficients decline strongly with the lengthening horizon between Q0 and Q2, much less so for Q3 and Q4, for both the individual and group mean forecasts. Only for UR, where the correlations are high for reasons already noted, do these declines extend clearly through the entire target range (Q0-Q4). The dispersion of the $\bar{r}^2$ coefficients across the individuals declines as the distance to the target quarter increases, except for UR where the opposite happens (column 8).

IV. Summary and Interpretations of Findings

The results of the study support the following statements:

1. The group mean forecasts from a series of surveys on the average over time more accurate than most of the corresponding sets of individual predictions. This is a strong conclusion, which applies to all variables and predictive horizons covered and is consistent with evidence for different periods and from other studies. It is based on an intensive analysis of a large collection of authentic macroeconomic forecasts, in two forms: (1) individual-to-group RMSE ratios M_i/M_{gi}, which turn out to be predominantly larger than 1.0, and (2) distributions of summary measures of accuracy, in which the series of the overall mean predictions M_g place better than half or more of the individuals.

2. The minorities that did succeed in outperforming the group averages vary in size across the variables (from 20% for GNP to 33% for CBI) and, particularly, across the horizons Q0-Q4 (e.g., 11-37% for IPD,

8-42% for UR). In each of the thirty categories combining specific variables and target quarters, most of the forecasters show RMS errors exceeding those of the strictly comparable group mean forecasts, and in most of the categories these majorities are large. The M_i/M_{gi} ratios average 1.1 and cluster between 0.9 and 1.4.

3. Rank correlations among the respondents according to the same ratios are positive for all variables and target quarters, and they are statistically significant in most cases by the conventional tests. For this result to obtain, a moderate degree of consistency must have existed in the relative performance of a sufficient number of the survey members. It is still true, as earlier reports also indicate, that no single forecaster has been observed to earn a long record of superior overall accuracy,[13] and indeed nothing in the present study would encourage us to expect any individual to reach this elusive goal. But a small number of the more regular participants in the ASA-NBER surveys did perform better in most respects than the composite forecasts from the same surveys.

4. To go beyond the observations in point 3 above, a further study of the characteristics, methods, and results of the forecasters with the best records will be needed. To mention just one question of interest, it remains to be seen whether weighted combinations of selected forecasts from this subgroup would yield significantly large and persistent gains in accuracy, but our results do not rule out this possibility. It

[13] See, e.g., Zarnowitz, 1967, pp. 123-132; and McNees, 1979, pp. 4-17.

seems more doubtful that weighting could be applied with much benefit directly to large numbers of forecasts from the surveys.[14]

5. Absolute measures of error depend strongly on the characteristics of the predicted variables and vary accordingly, in contrast to the standardized M_i/M_{gi} ratios. For example, relatively smooth series such as the unemployment rate and growth in nominal GNP are easier to predict and are in fact much better predicted than the more volatile series such as growth in real GNP and the IPD inflation, as indicated by comparisons of average size and variability of forecast errors and realizations.

6. The overall composite forecasts M_g have RMSE's that are for almost all categories smaller than the medians, and indeed often close to the lower quartiles, of the distributions of the RMSE's for the corresponding individual forecasts. Also, the correlations of predicted with actual values $(\bar{r}^2)$ are typically higher for M_g than for most of the individuals, frequently by substantial margins. These results are apparently unrelated to the differential characteristics of the variables covered.

7. The location and dispersion statistics for the distributions of the RMSE and $\bar{r}^2$ measures display much diversity but also some apparent regularities. The medians tend to be smaller than the means, suggesting some positive skewness in the RMSE distributions. While the standard deviations of the individual RMS errors and $\bar{r}^2$ coefficients vary greatly across the different variables, the coefficients of variation do not.

---

[14]Under circumstances that are not infrequently encountered in practice, equal weighting schemes have been found to yield more accurate composite forecasts than differential weighting schemes derived by least squares; see Einhorn and Hogarth, 1975.

8. There is a general tendency for the errors to increase in absolute size with the time distance to the target quarter, but by decreasing margins. Also, correlations between predictions and realizations typically decline as the target period recedes into the future, but again more so for the nearest than for the more distant quarters. The relative dispersion measures tend to decrease with the predictive horizon for the RMSE's and rise for the correlation statistics, while the absolute dispersion measures show no common patterns of change.

References

Christ, Carl F. 1975. "Judging the Performance of Econometric Models of the U. S. Economy," International Economic Review 16, no. 1 (February): 54-74.

Einhorn, Hillel, and Hogarth, Robin. 1975. "Unit Weighting Schemes for Decision Making," Organizational Behavior and Human Performance 13, no. 2 (April): 171-192.

Fair, Ray C. 1974. "An Evaluation of a Short-Run Forecasting Model," International Economic Review 15, no. 2 (June): 285-303.

McNees, Stephen K. 1973. "The Predictive Accuracy of Econometric Forecasts," New England Economic Review (September/October): 3-27.

McNees, Stephen K. 1975. "An Evaluation of Economic Forecasts," New England Economic Review (November/December): 25-29.

McNees, Stephen K. 1976. "An Evaluation of Economic Forecasts: Extension and Update," New England Economic Review (September/October): 30-44.

McNees, Stephen K. 1979. "The Forecasting Record of the 1970s," New England Economic Review (September/October): 33-53.

Moore, Geoffrey H. 1969. "Forecasting Short Term Economic Change," Journal of the American Statistical Association 64, no. 325 (March): 1-22.

Moore, Geoffrey H. 1977. "Lessons of the 1973-1976 Recession and Recovery," in William Fellner, ed., Contemporary Economic Problems 1977.

Su, Vincent, and Su, Jospehine. 1975. "An Evaluation of the ASA/NBER Business Outlook Survey Forecasts," Explorations in Economic Research 2, no. 4 (Fall): 588-618.

Zarnowitz, Victor. 1967. An Appraisal of Short-Term Economic Forecasts. New York: National Bureau of Economic Research.

Zarnowitz, Victor. 1972. "Forecasting Economic Conditions: The Record and the Prospect," in V. Zarnowitz, ed., The Business Cycle Today. New York: National Bureau of Economic Research: 183-239.

Zarnowitz, Victor. 1979. "An Analysis of Annual and Multiperiod Quarterly Forecasts of Aggregate Income, Output, and the Price Level," Journal of Business 52, no. 1 (January): 1-33.

Zarnowitz, Victor. (1982). "Expectations and Forecasts from Business Outlook Surveys," NBER Working Paper No. 845 (January).

Zarnowitz, Victor. 1983. "Rational Expectations and Macroeconomic Forecasts," Journal of Macroecnomics (forthcoming).