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

Volume Author/Editor: Victor Zarnowitz

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-97890-7

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

Conference Date: n/a

Publication Date: January 1992

Chapter Title: The Accuracy of Individual and Group Forecasts

Chapter Author: Victor Zarnowitz

Chapter URL: http://www.nber.org/chapters/c10386

Chapter pages in book: (p. 444 - 461)

15 The Accuracy of Individual and Group Forecasts

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-sectional and distributional aspects of the data: differences among the individual 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 chapter 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 econ-

Reprinted from the *Journal of Forecasting* 3 (Jan.–March 1984): 11–26.

Financial support from the U.S. National Science Foundation and aid of the Graduate School of Business of the University of Chicago and NBER are gratefully acknowledged. The author is greatly indebted to Louis Lambros for helpful suggestions and efficient research assistance. Earlier versions of this paper were presented in workshops at the University of Chicago and Stanford University, at the annual meeting of the American Association for the Advancement of Science in Washington, D.C., on 8 January 1982, and at the Third International Symposium on Forecasting in Philadelphia on 6 June 1983. The paper has benefited from comments of the participants, of many colleagues, and anonymous referees.

1. An early study which dealt with certain characteristics of the relation between aggregate and individual forecasts is Zarnowitz 1967, pp. 123–26. A more recent analysis of disaggregated data from surveys of inflation forecasts is Figlewski and Wachtel 1981.

omy, 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, government agencies, and academic and research 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 to 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

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

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 equation (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 marginal or "intraforecast" 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

$$(2)\qquad E_{t+j} = P_{t+j} - A_{t+j}, \qquad j = 0, 1, \ldots, 4,$$

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 1969a, 1977c; Zarnowitz 1972b, 1979; Fair 1974; Christ 1975; McNees 1975, 1976; Su and Su 1975.

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

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, and 1979, p. 6, and McNees 1973, pp. 7–10.

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.[5]

15.1 Measures of Relative Accuracy and Consistency

The root mean square error (RMSE) of the *i*th individual's set of predictions can be written as

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

for any variable and forecast horizon.[6] Here $\{N_i\}$ is the set of the target periods of the *i*th forecasts, and 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 equation (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 (g_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

(4) $$M_{gi} = \left(\frac{1}{n_i} \sum_{t \in N_i} \varepsilon_{gt}^2\right)^{1/2}.$$

Ratios of RMSEs, M_i/M_{gi} provide convenient measures of the 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. For any

5. For a report on tests of bias or rationality, see Zarnowitz 1983.

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).

target category, the most accurate individuals will have $M_i < M_{gi}$, and the proportion of such forecasts can vary over a wide range. However, if all forecasters participated in every survey (no gaps), then the average of the corresponding RMSEs across the individuals could not exceed the group RMSE. That is,

$$\frac{1}{m}\sum_{i=1}^{m} M_i \geq M_g$$

for a group g consisting of m regular participants, provided that

$$\varepsilon_{gt} = \frac{1}{m}\sum_{i=1}^{m} \varepsilon_{it}$$

(which is easily demonstrated for simple level forecasts).[8]

It follows that a group average forecast has a built-in advantage vis-à-vis individual forecasts: a strategy of random selection among the latter will involve a greater risk of error than that of using the average. However, this by no means precludes that some individuals may be superior forecasters whose performance has been above-average over time due to better skills or methods. It is indeed of particular interest to compare the individuals with survey averages which represent reasonably accessible and efficient forecasting benchmarks. Data on the median forecasts from the ASA-NBER surveys (which resemble the overall group means used later in this study) 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 widely used predictions for the U.S. economy.

Contemporaneous expectations for a given target may be distributed more or less symmetrically about their mean, but over time the individual's positions within these distributions are likely to fluctuate. For most people, most of the time, the predictive record may be spotty, with but transitory spells of relatively high accuracy. A series of group averages is helped by offsetting errors, in particular by the cancellation of individual errors of opposite sign.[9] But the gains from combining generally good professional forecasts into a set of composite predictions (e.g., the group means) depend on the existence of

8. This is a special case of Minkowski's inequality. See Hardy, Littlewood, and Pólya 1964, theorem 24, pp. 30–31. I am indebted to Michael Rothschild and Edward George for a clarification of this point.

M_i is here defined as $\left(\frac{1}{n}\sum_{t=1}^{n} \varepsilon_{it}^2\right)^{1/2}$ and M_g as $\left[\frac{1}{n}\sum_{t=1}^{n}\left(\frac{1}{m}\sum_{i=1}^{m} \varepsilon_{it}\right)^2\right]^{1/2}$.

Average M_i will be equal to M_g only when the forecasts and hence their errors ε_i, are proportional.

9. See Zarnowitz 1967, p. 125, for a related discussion. Stekler and Thomas (1980) use "penalized MSE" measures involving sums $\Sigma_i(e_{it})^2$ rather than $(\Sigma_i e_{it})^2$ to determine the importance of offsetting errors in component forecasts (e) for the evaluation of aggregate forecasts.

independent elements in the individual forecasts. It is these facts that will be found of primary importance in our comparative analysis.

15.1.1 The M_i/M_{gi} Ratios: Individuals versus Group Averages

Inspection of graphs for 30 distributions of ratios of RMSEs, M_i/M_{gi} (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 figure 15.1, 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. If the forecasts ε_{it} were independent normal variables, then ε_{it}^2 and their sums would have chi-square distributions. This may help explain the skewness of the distributions of the RMSE ratios shown in figure 15.1.

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 $\bar{M}$) are all located to the right of the unity (broken vertical) lines, as they would have to be if the inequality referred to before (see text and n. 8) applied strictly. The histograms tend to get tighter and also, often, less skewed for the more distant quarters.[10]

Table 15.1 shows that the mean ratios are remarkably close: when rounded, all but 9 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 decrease strongly with the distance to the target quarter, from Q0 to Q3.[11] 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%–42%; for CBI it is only 29%–38%. Averaged across Q0–Q4, the figures fall between 20% for GNP and 33% for CBI (see the last section of table 15.1).

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

10. See Zarnowitz 1982b, pp. 16–18, for a chart showing all the graphs discussed in the text above.

11. The series for Q0, Q1, Q2, and Q3 start in 1968:4, 1969:1, 1969:2, and 1969:3, respectively, and extend to 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 with those for Q3 probably reflect the drop in survey coverage.

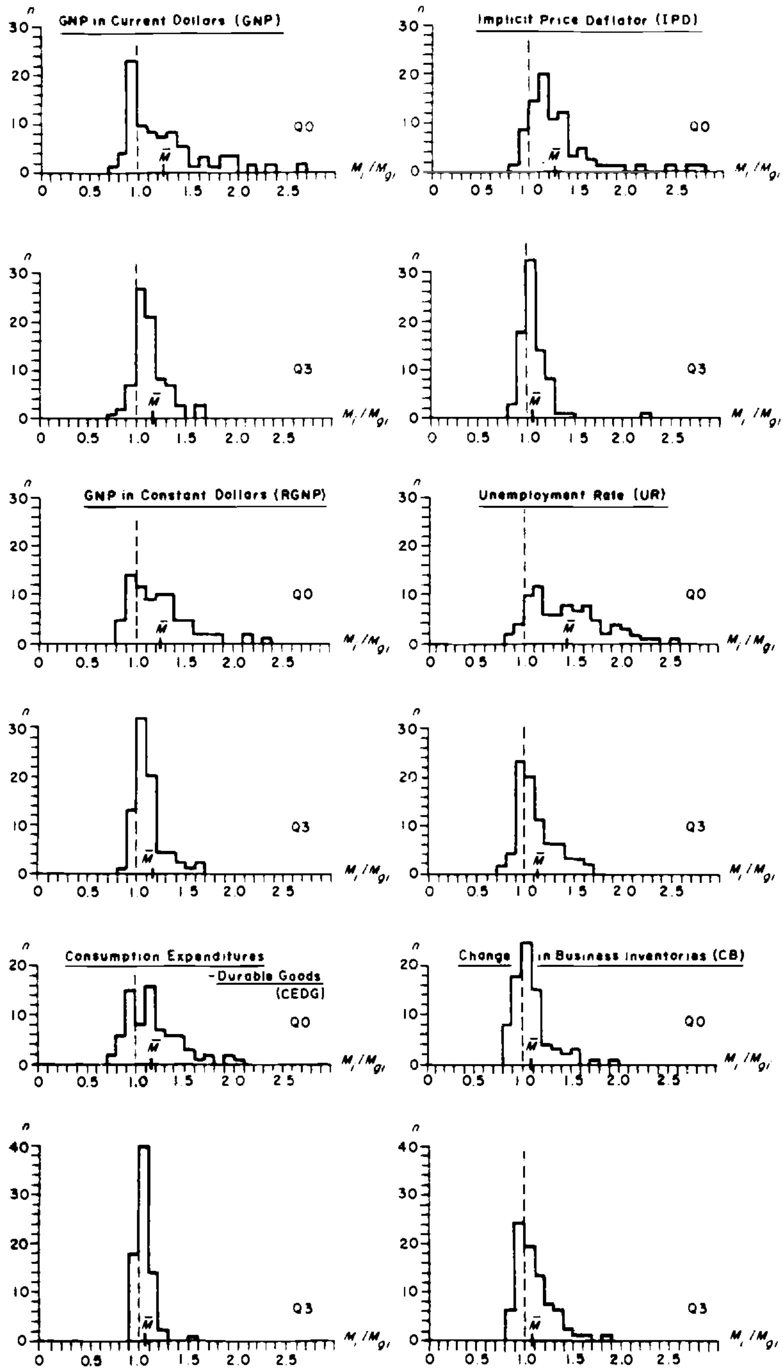

Fig. 15.1 Seventy-nine individual forecasts of multiperiod changes in six aggregate variables, comparison with group mean forecasts, 1968–79

Table 15.1 Means and Standard Deviations of the M_i/M_{gi} Ratios, by Variable and Target Quarter, 1968–79

	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	0.38	0.36	0.31	0.36	0.28	0.20
Q1	0.21	0.23	0.18	0.24	0.18	0.21
Q2	0.18	0.16	0.15	0.19	0.11	0.19
Q3	0.17	0.18	0.15	0.18	0.09	0.18
Q4	0.24	0.20	0.19	0.19	0.10	0.19
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

Source: Based on quarterly ASA-NBER business outlook surveys, 1968:4–1979:1.
Note: On coverage and symbols used, see text and figure 15.1.

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.

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.[12] 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 $j = 0, 1, \ldots, 4$).

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 6-month and 12-month forecasts of in-

12. 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 15.1, col. 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).

dustrial 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.[13]

15.1.2 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 15.2, both across the variables for each target quarter and across target quarters for each variable. All the correlations are positive and in general they appear to be significantly so (see note to table 15.2). 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 $\bar{\rho}$ is 0.74 in this case. For variables that are not so 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 A 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 ρs for Q0–Q1 average 0.61, those for Q1–Q2, Q2–Q3, and Q3–Q4 average 0.52–0.55 (see pt. B of table 15.2). For nonadjoining 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

13. See Zarnowitz 1982b, 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. 123–26) tell much the same story.

Table 15.2 **Rank Correlations among Participants in ASA-NBER Surveys according to Ratios of Individual to Group RMSEs. Multiperiod Predictions for Six Aggregate Variables, 1968–79**

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

	GNP	IPD	RGNP	UR	CEDG	CBI		GNP	IPD	RGNP	UR	CEDG	CBI
Target quarter Q0							*Target quarter Q1*						
GNP	1.00						GNP	1.00					
IPD	0.57	1.00					IPD	0.23	1.00				
RGNP	0.83	0.65	1.00				RNGP	0.77	0.48	1.00			
UR	0.42	0.43	0.39	1.00			UR	0.20	0.39	0.40	1.00		
CEDG	0.69	0.56	0.69	0.50	1.00		CEDG	0.43	0.20	0.50	0.30	1.00	
CBI	0.40	0.41	0.36	0.21	0.42	1.00	CBI	0.39	0.31	0.38	0.22	0.38	1.00
Target quarter Q2							*Target quarter Q3*						
GNP	1.00						GNP	1.00					
IPD	0.35	1.00					IPD	0.27	1.00				
RGNP	0.66	0.48	1.00				RGNP	0.69	0.33	1.00			
UR	0.23	0.48	0.27	1.00			UR	0.41	0.44	0.49	1.00		
CEDG	0.27	0.15	0.21	0.12	1.00		CEDG	0.21	0.05	0.14	0.19	1.00	
CBI	0.41	0.43	0.32	0.36	0.31	1.00	CBI	0.28	0.42	0.15	0.37	0.03	1.00
Target quarter Q4							*Average Q0–Q4*						
GNP	1.00						GNP	1.00					
IPD	0.44	1.00					IPD	0.37	1.00				
RGNP	0.76	0.51	1.00				RGNP	0.74	0.49	1.00			
UR	0.39	0.49	0.31	1.00			UR	0.33	0.45	0.35	1.00		
CEDG	0.18	0.17	0.17	0.06	1.00		CEDG	0.36	0.23	0.34	0.23	1.00	
CBI	0.49	0.36	0.38	0.19	0.21	1.00	CBI	0.32	0.39	0.32	0.27	0.27	1.00

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

	Q0	Q1	Q2	Q3	Q4		Q0	Q1	Q2	Q3	Q4
GNP in current dollars (GNP)						*Implicit price deflator (IPD)*					
Q0	1.00					Q0	1.00				
Q1	0.51	1.00				Q1	0.55	1.00			
Q2	0.18	0.19	1.00			Q2	0.45	0.68	1.00		
Q3	0.14	0.18	0.40	1.00		Q3	0.41	0.54	0.60	1.00	
Q4	0.50	0.32	0.40	0.47	1.00	Q4	0.39	0.52	0.51	0.62	1.00
GNP in constant dollars (RGNP)						*Unemployment rate (UR)*					
Q0	1.00					Q0	1.00				
Q1	0.57	1.00				Q1	0.64	1.00			
Q2	0.33	0.47	1.00			Q2	0.38	0.78	1.00		
Q3	0.05	0.19	0.48	1.00		Q3	0.32	0.62	0.85	1.00	
Q4	0.38	0.19	0.33	0.33	1.00	Q4	0.27	0.53	0.75	0.92	1.00
Consumer expenditures for durable goods (CEDG)						*Change in business inventories (CBI)*					
Q0	1.00					Q0	1.00				
Q1	0.66	1.00				Q1	0.70	1.00			
Q2	0.53	0.43	1.00			Q2	0.63	0.76	1.00		
Q3	0.12	0.07	0.14	1.00		Q3	0.51	0.56	0.81	1.00	
Q4	0.14	0.00	0.04	0.08	1.00	Q4	0.57	0.62	0.63	0.76	1.00

Note: For rankings without ties, the variance of ρ equals $n/(n - 1)$ (Kendall 1948, p. 46). For $n = 75$, therefore, the standard error $s_\rho = 1/\surd(0.74) = 0.1162$ (for $n = 79$, $s_\rho = 0.1124$; 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.

[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 acronyms for the variables are identified in part B of the table. 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.

[b]These measures refer to the sample covered in fig. 15.1:79 individuals for each of the variables except CEDG (80). The rank correlation coefficients are Spearman's ρ.

CEDG, GNP, and RGNP (0.27–0.33) and high for IPD, UR, and CBI (0.55–0.66).

15.2 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).[14]

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 30, 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 n. 11).

15.2.1 Overall Accuracy

For each of the six variables, the means of the individual RMSEs taken across the target quarters Q0, . . . , Q4 exceed the corresponding RMSEs for the overall group mean forecasts. The ratios of the summary statistics of error (entries in col. 2 of table 15.3 divided by those in col. 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 for the individuals vary between 0.14 and 0.28, except for the unemployment rate, a

14. See Zarnowitz 1983 for more numerical detail on the forecast samples from the ASA-NBER surveys.

Table 15.3 **Selected Overall Accuracy Statistics for Individual and Group Mean Forecasts, Six Variables, 1968–79**

	Individual Forecasts[a]			Group Mean Forecasts[b]			Actual Values[c]		
Variable	ME (1)	RMSE (2)	$\bar{r}^2$ (2)	ME (4)	RMSE (5)	$\bar{r}^2$ (6)	Mean (7)	S.D. (8)	RMSV (9)
GNP	−0.11	1.00	0.22	−0.13	0.88	0.29	2.18	1.09	2.93
IPD	−0.39	0.78	0.21	−0.38	0.67	0.27	1.50	0.67	1.64
RGNP	0.28	1.21	0.28	0.25	1.05	0.35	0.68	1.24	1.41
UR	−0.14	0.67	0.66	−0.10	0.62	0.69	5.78	1.68	6.02
CEDG	−0.28	4.04	0.14	−0.36	3.68	0.18	2.25	3.97	4.57
CBI	−1.72	10.00	0.27	−1.77	9.57	0.44	9.19	10.87	14.23

[a] These measures refer to the sample covered in fig. 15.1 (75 individuals forecast CEDG; 79 individuals forecast 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.

[b] 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.

[c] For the definition of actual values, see text. For the nominal and real gross national product (GNP and RGNP), IPD, and CEDG, the measures refer to percentage changes; for UP and CBI, they refer to levels. S.D. = standard deviation (corresponding to the means in col. 7): RMSV = root mean square value computed as $\sqrt{[(\text{mean})^2 + (\text{S.D.})^2]}$.

relatively smooth level series, where the $\bar{r}^2$ is 0.66 (col. 3). They are 60%–80% lower than their counterparts for the overall group means (col. 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. cols. 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 orders of magnitude 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 15.3.

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. cols. 1 and 4 with col. 7). More telling, the RMSEs are less than the corresponding root mean square values of the target series, again in most cases by large margins (cf. cols. 2 and 5 with col. 9). The RMSEs for the group mean forecasts are also generally less than the standard deviations of the actual values (cols. 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.

15.2.2 Characteristics of the Distributions

The medians of the RMSEs for the individual forecasts are with few exceptions lower than the means, but by relatively small margins (see table 15.4, cols. 1 and 4). This indicates a weak tendency for these distributions to be skewed to the right, that is, toward the large RMSEs.

In virtually all instances, the averages of the individual RMSEs exceed the RMSEs for the corresponding group mean forecasts (compare the entries in cols. 1 and 4 with their counterparts in col. 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 RMSEs (cf. cols. 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 $M_i < M_{gi}$ is about 26% (table 15.1).

The more distant the target quarter, the larger tend to be the prediction errors, as demonstrated by the increases from Q0 to Q4 of the entries in columns 1 and 3–6 of table 15.4. 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 versus Q3. In short, these measures suggest that the RMSEs 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, . . .).[15]

15. 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

The absolute dispersion measures (standard deviations in col. 2 and interquartile ranges implied by cols. 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, namely, the coefficients of variation S.D./M (ratios of entries in col. 2 to those in col. 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 0.23 to 0.26, except for CEDG and CBI, where they are 0.19 and 0.31, respectively).

The group mean forecasts have tracked the actual changes better than the average individual forecasts: the correlation measures in column 9 of table 15.4 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 (col. 8).

15.3 Summary and Interpretations of Findings

The results of the study support the following statements:

1. The group mean forecasts from a series of surveys are 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 30 categories combining specific variables and target quarters, most of the forecasters show RMSEs 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.

quarters ahead, without any systematic upward drift (Zarnowitz 1967, pp. 64–72, 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 controlled data.

Table 15.4 **RMSEs and Correlations between Predicted and Actual Values, Selected Distributional Statistics by Variable and Target Quarter, Individual and Group Mean Forecasts, 1968–79**

	RMSEs						Squared Correlations ($\bar{r}^2$)		
	Individual Forecasts						Individual Forecasts		
Quarter Predicted	*M* (1)	S.D. (2)	LQ (3)	Me (4)	UQ (5)	Group Mean M_g (6)	*M* (7)	S.D. (8)	Group Mean, M_g (9)
GNP in current dollars (GNP)									
Q0	0.77	0.27	0.58	0.68	0.92	0.66	0.49	0.22	0.63
Q1	0.95	0.26	0.78	0.92	1.05	0.86	0.24	0.16	0.31
Q2	1.06	0.19	0.96	1.07	1.15	0.96	0.13	0.12	0.18
Q3	1.10	0.26	0.92	1.10	1.23	0.98	0.12	0.13	0.16
Q4	1.12	0.28	0.94	1.08	1.22	0.94	0.12	0.15	0.18
Implicit price deflator (IPD)									
Q0	0.55	0.16	0.45	0.49	0.60	0.42	0.45	0.20	0.64
Q1	0.69	0.16	0.58	0.66	0.77	0.59	0.28	0.18	0.35
Q2	0.79	0.16	0.69	0.78	0.87	0.70	0.14	0.12	0.17
Q3	0.88	0.19	0.78	0.86	0.95	0.77	0.10	0.10	0.12
Q4	0.98	0.21	0.86	0.94	1.09	0.88	0.10	0.10	0.08
GNP in constant dollars (RGNP)									
Q0	0.85	0.28	0.67	0.78	0.96	0.70	0.60	0.18	0.75
Q1	1.09	0.28	0.91	1.03	1.26	0.95	0.38	0.17	0.48

Q2	1.24	0.22	1.13	1.25	1.37	1.12	0.18	0.13	0.25
Q3	1.39	0.25	1.22	1.36	1.53	1.23	0.10	0.11	0.10
Q4	1.46	0.31	1.25	1.39	1.69	1.23	0.12	0.15	0.16
Unemployment rate (UR)									
Q0	0.22	0.06	0.17	0.21	0.26	0.16	0.97	0.02	0.99
Q1	0.46	0.11	0.38	0.44	0.51	0.41	0.86	0.06	0.91
Q2	0.71	0.17	0.60	0.67	0.81	0.65	0.68	0.12	0.75
Q3	0.94	0.23	0.78	0.91	1.09	0.88	0.48	0.17	0.53
Q4	1.04	0.24	0.88	1.00	1.19	0.98	0.32	0.19	0.27
Consumer expenditures for durable goods (CEDG)									
Q0	3.37	0.96	2.64	3.10	3.85	2.87	0.40	0.25	0.63
Q1	4.16	0.78	3.79	4.13	4.60	3.77	0.11	0.15	0.13
Q2	4.24	0.71	3.82	4.22	4.68	4.04	0.05	0.07	0.01
Q3	4.44	0.65	4.12	4.41	4.78	4.09	0.05	0.07	0.01
Q4	3.98	0.69	3.47	3.99	4.47	3.64	0.08	0.08	0.13
Change in business inventories (CBI)									
Q0	8.21	2.65	6.70	8.10	9.61	8.07	0.36	0.21	0.55
Q1	9.17	3.06	7.06	8.89	11.86	9.11	0.31	0.22	0.51
Q2	10.42	3.36	8.22	10.08	12.87	9.79	0.25	0.19	0.41
Q3	10.99	3.32	9.16	11.12	13.10	10.08	0.20	0.18	0.40
Q4	11.22	3.00	9.13	11.14	13.12	10.80	0.21	0.17	0.35

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 n. 11). M = mean; S.D. = standard deviation; Me = median; 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.

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,[16] 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, 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.[17] It seems more doubtful that weighting could be applied with much benefit directly to large numbers of forecasts from the surveys.[18]

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 RMSEs that are for almost all categories smaller than the medians, and indeed often close to the lower quartiles, of the distributions of the RMSEs 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. Although the standard deviations of the individual RMSEs and $\bar{r}^2$ coefficients vary greatly across the different variables, the coefficients of variation do not.

16. See, for instance, Zarnowitz 1967, pp. 123–32 and McNees 1979, pp. 4–17.

17. On weighted combinations of forecasts, see Bates and Granger 1969 and Granger and Newbold 1977, pp. 269–78.

18. 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 RMSEs and rise for the correlation statistics, whereas the absolute dispersion measures show no common patterns of change.

Forecast makers and users may draw the following conclusions from these findings and some related results in the literature:

1. It is advantageous for the experts to consider various methods and sources of prediction, including the recent evolution of the "consensus" of expectations. Good practitioners absorb a great deal of common information, which tends to both improve their forecasts and make them similar. At any point in time, luck may count as much as skill in ranking the forecasters, but on the average over time those who use better models, techniques, and judgments are likely to score significantly above average. It is these individuals or teams, working in large measure independently, that contribute to any successful forecasting which deserves attention, whereas mere opinions of the "follow-the-leader" or similar types do not.[19]

2. Just as there are gains to the forecast makers from combining different, relevant, and complementary approaches, so there are gains to the forecast users from combining predictions from different sources, provided that the latter are sufficiently independent. This suggests that decision and policy makers do well to consult the leading surveys of economic and business forecasters as well as any of their favored individual sources. Where available, measures of dispersion of the forecasts are worth monitoring along the averages.

3. The survey averages include some econometric and many judgmental forecasts. They are neither better nor worse than the predictions from the well-known econometric service bureaus in any systematic sense; the detailed comparisons differ by variable, span, and target period. Forecasts from the major econometric model services generally benefit from judgmental adjustments. Although very influential, they are themselves also influenced by other predictions originating in large industrial corporations, financial institutions, and government agencies. Such cross-effects blur somewhat the distinctions between the forecasts involved, but they stop far short of eliminating them. The interactions persist because they are deemed useful.

19. Thus I do not believe that the results of this study or other evidence "support the wisdom of changing forecasts to go along with the crowd" (Silk 1983). First, the "crowd" itself is a select one here and, second, a sizable minority of individuals was shown to have produced forecasts superior to the overall averages.