NBER WORKING PAPER SERIES

RATIONAL EXPECTATIONS AND MACROECONOMIC FORECASTS

Victor Zarnowitz

Working Paper No. 1070

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138

January 1983

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, and to Paul Flignor and Scott Katzer for help with computations. 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 #1070
January 1983

RATIONAL EXPECTATIONS AND MACROECONOMIC FORECASTS

Abstract

This paper presents extensive results from testing for bias and serially correlated errors in a large collection of quarterly multiperiod predictions from surveys conducted since 1968 by the National Bureau of Economic Research and the American Statistical Association. The tests of the joint null hypothesis that the regressions of actual on predicted values have zero intercepts and unitary slope coefficients are very unfavorable to the expectations of inflation, but they show the forecasts of several other variables in a generally much better light. There have been strong tendencies for the forecasters in this period to underestimate inflation and overestimate real growth. Considerable attention is given to the effects of the sample size--the issue of the power of the tests--and also to the extent and role of autocorrelations among the residual errors from these regressions.

Rationality in the sense of efficient use of relevant information implies the absence of systematic elements in series of errors from the forecaster's own predictions, measured strictly in the form in which such errors could have been known at the time of the forecast. The frequencies of significant autocorrelations among errors so measured vary greatly across the forecasts for different variables, being very high for inflation, high for inventory investment and the unemployment rate, and much lower for most of the predictions of the other variables covered (rates of change in nominal and real GNP and expenditures on consumer durables). The corresponding tests for the group mean forecasts show much less evidence of serially correlated ex ante errors, except for inflation.

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

(312) 962-7130

I. Questions and Data

On Economics of Expectations and Surveys of Forecasts

Much effort was spent in recent years on collecting and processing data from periodic surveys of intentions, plans, or predictions of various groups: consumers, corporate managers, business and financial analysts, economists. This work was motivated mainly by the prospect of obtaining useful tools for practical forecasting, but it is increasingly recognized that the data can have important analytical uses for measurement and study of economic expectations.

Recent theorizing about expectations concentrates on market prices and rewards that motivate people to use all information that can be acquired cost-effectively. The rational expectations hypothesis assumes that a sufficiently large number of agents know "how the world works," that is, recognize the structure of their environment and efficiently process all available and pertinent data. It is the so formed expectations that are decisive for what transpires in the market place, and they are reflected in the equilibrating behavior of prices and other endogenous variables (Muth, 1961; Poole, 1976). Prices in a market may incorporate all information that matters, even though price expectations of many, perhaps even most, traders do not meet the rationality criterion.[1]

---

[1] For this to happen, all that is needed is that some resourceful participants have their way in eliminating the unexploited profit opportunities in the given market. Those who succeed relatively often tend to reap gains: the competitive game of economic prediction cannot be comprehended by treating expectations as if they were simple-valued and universally shared. Thus it is important to distinguish between <u>individual</u> and <u>market</u> expectations. For an early argument that rational market reactions may coexist with a large amount of individual "irrationality," see Becker, 1962.

However, under uncertainty and in areas of the economy other than the competitive auction markets, quantity signals may be as important as price signals. Economic agents are presumably most interested in local variables relating closely to their own activities, but aggregate measures such as real GNP growth, inflation, unemployment, sensitive cyclical indicators, changes in money and credit, interest rates, and exchange rates are also widely monitored and selectively used. For most of the macrovariables, market expectations are nonexistent or unobservable, but it is evident that numerous predictions are being regularly made and used throughout the economy. Macropredictions serve as important inputs to micropredictions.

Not surprisingly, professional business analysts and economists produce the bulk of the macroeconomic predictions, both for public and internal uses, and many of them participate in periodic business outlook surveys. It might be argued that these are forecasts of people who study the economy (experts), which are quite unlike the expectations of those who act in the economy (agents). On the one hand, the experts are usually credited with more knowledge of the economy at large than the agents have. On the other hand, the experts are often charged with being less strongly motivated to predict optimally than the agents who are seen as having more at stake.

In practice, the distinction between agents and experts is at this point very blurred. Macroeconomic forecasters who sell their services to governmental and corporate decision makers and often compete as well in the market for public attention are treated as "experts" but they are certainly also "agents" in their own rights. Indeed, many of them are influential agents who have passed critical market tests, as certified by their positions and by the rewards their forecasts and advice earn them in the business world. It can be presumed that, in general, they do have incentives to perform well and strive to do so.

Consistent with this view, it is appropriate that the results of business outlook surveys have received alternative interpretations in the literature. They are treated either as agents' expectations, e.g., in tests of whether they conform to the hypotheses of rational or adaptive expectations, or as experts' forecasts, e.g., in comparisons with predictions from particular econometric models.[2] This paper will adopt the first of these perspectives.

An ideal survey would use a large, properly constructed random sample to insure that the respondents represent well the universe of those whose expectations count, and a system of rewards and penalties to insure that they have a stake in their responses. Of course, the ideal surveys do not exist and the actual ones may be far from ideal. If a survey yields inferior or biased predictions, it is possible that carelessness, poor information, or other failings of particular respondents are to blame, which should not be generalized. The evidence may be distorted and the results misinterpreted because of reporting errors, outliers, undue reliance on averages from small samples, spotty participation, or limited time coverage. But detailed knowledge of, and attention to, the data can go far to safeguard against such pitfalls. This work should benefit from the author's direct involvement with the management of the surveys to be discussed.

Tests of Rationality

Rational expectations *sensu stricto* satisfy

$$E(y_t^* | I_{t-1}) = (y_t), \quad t = 1, \ldots, n , \tag{1}$$

---

[2]For examples and further references, see Theil, 1965; Mincer, 1969; Mincer and Zarnowitz, 1969; Zarnowitz, 1974, and 1979; McNees, 1978; Nelson, 1975; Carlson, 1977; Wachtel, 1977; Pearce, 1979; Figlewski and Wachtel, 1981.

where y_t^* is the one-period-ahead prediction of the variable y_t; E is the expected value operator; and I_{t-1} is the set of all information (data and models) on which y_t^* was conditioned at the time it was made. All attempts to apply this abstract formula confront a dilemma. To determine whether the predictions y^* are rational in the sense of (1), I_{t-1} must be specified, but as a rule the outside observer has no way of knowing what this set contains. (Indeed, even the source of a particular value of y^* would probably often find it difficult to define the contents of I_{t-1} clearly and exhaustively.)[3]

If adequate data on y^* are available, it is possible to test one implication of rationality, namely lack of bias

(2) $$E(y_t^* - y_t) = 0 .$$

To this end, the regression

(3) $$y_t = a + by_t^* + u_t$$

is estimated to verify or falsify the joint hypothesis that a and b are not statistically different from 0 and 1, respectively. However, this is a weak test, since rational expectations imply efficient use of pertinent information, not just unbiasedness. And unbiased predictions may still be far from optimal or even accurate.

---

[3]Consider as an example the much studied short-term expectations of inflation: what is known about their determinants? There are the dominant hypotheses of economic theory. But economists do not agree on all the important features of their models, and insofar as their models contradict each other they surely cannot all be properly specified. It is difficult to accept the notion that the representative agent is free of the limitations of knowledge that are evident in experts' analysis of the economy. But consequences of incomplete information or deficient knowledge may be mistaken for departure from rational expectations (Zarnowitz, 1982a).

The advantage of testing H_0: $(a, b) = (0, 1)$ is that no specification is needed of what information the forecasters could and should have used, and how. But it is possible to use a considerably stronger criterion of rationality without getting involved in difficult and to some extent inevitably arbitrary assumptions about the plausible data and models constituting the information sets in question. For any variable, an important part of the set I_{t-1} is made of past errors made by the forecaster and known (or at least knowable) to him or her at the time of the forecast. The testable requirement here is that there be no significant autocorrelation among such errors, i.e., that the predictions be essentially free of systematic error components that could have been detected and corrected on a current basis.

In this study, the tests of bias and autocorrelation of errors are applied to a large number of time series of multiperiod predictions for six selected macroeconomic variables. The data, described below, are believed to represent well the contemporary "state of the art" in professional forecasting of business conditions. Problems of how to measure the predictive errors and how to estimate the parameters in question are best discussed in the context of the actual data used.

Sources of Evidence and Scope of Study

Owing to the efforts of the National Bureau of Economic Research, in collaboration with the American Statistical Association, a large amount of information has been assembled on the record of forecasting changes in the U. S. economy. Each quarter, the NBER examines the results of a questionnaire mailed by the ASA.[4] The survey reaches a broadly based and diversified group

---

[4]For the quarterly reports on each survey, see NBER Explorations in Economic Research (through 1977) and NBER Reporter (since 1978). The corresponding ASA reports have appeared in the American Statistician and

of persons who are regularly engaged in the analysis of current and prospective business conditions. Most of the respondents are from the world of corporate business and finance but academic institutions, government, consulting firms, trade associations, and labor unions are also represented. The format of the survey remained unchanged from its inception in 1968:4 through 1981:2, with forecasts covering on each occasion the current and the next four quarters, for eleven time series representing the principal measures of national output, income, consumption, investment, the price level, and unemployment.[5]

Past work on the survey data has concentrated on summary measures (mainly group medians or means, in some cases standard deviations), whereas this paper is part of a comprehensive study of forecasts by individual respondents in the NBER-ASA group. Further, unlike the many recent studies which consider only expectations of inflation, this report covers other important aggregative variables as well.

The body of the data consists of 42 consecutive surveys covering the period from 1968:4 through 1979:1. Altogether, the list of those who replied to any of the questionnaires includes 172 names (which are treated confidentially). However, many individuals responded only once or a few times, and

(since 1974) in AmStat News. The forecasts have been regularly published and frequently discussed in Economic Prospects, a report by the Commercial Credit Company (1972-73), and in Economic Outlook USA, a report by the Survey Research Center at the University of Michigan (since 1974). On the origin of the survey and the design of the questionnaire, see Zarnowitz, 1969a.

[5] In 1981 the coverage has been substantially extended. The surveys also have regularly collected unique data on the methods and assumptions used by the participants, and on the probabilities they attach to alternative prospects concerning changes in output and prices. For references to some evaluations of the overall results from the ASA-NBER surveys, see Zarnowitz, 1982c.

some decision had to be made on the minimum number of surveys that would qualify a participant for inclusion. It was set at 12, which still left as many as 79 individuals in the sample.

Four of the variables covered have strong upward trends, and it is not their levels that are of major interest but rather their rates of change which reflect their real growth and/or inflation. These are gross national product and consumer expenditures for durable goods, both in current dollars (GNP and CEDG); GNP in constant dollars (RGNP); and the GNP implicit price deflator (IPD). For these series, forecast errors are measured as differences, predicted minus actual percentage change.

The change in business inventories (CBI), a current-dollar series, is trendless, being already in first-difference form. The unemployment rate (UR) represents the percentage unemployed of the civilian labor force and is dominated by short-term, mainly cyclical movements, not a long-term trend. For these two variables, therefore, forecast errors are measured as differences, predicted level minus actual level.[6]

Including the group averages, about 400 quarterly time series of forecasts are available for each of the six variables (five series for as many target quarters per each of the 80 sources). The volume and quality of the data are such as to permit an intensive study of each of the various aspects of economic predictions.[7]

---

[6]See Zarnowitz, 1982c, for references to the treatment of level and change errors.

[7]Neglect of data problems explains why some survey evaluations yielded mixed and partly contradictory results of limited applicability. (A case in point is the series of surveys of economic forecasters conducted semiannually since 1947 by Joseph A. Livingston, a syndicated financial columnist. See Carlson, 1977; Pearce, 1979; and Figlewski and Wachtel, 1981). Several aspects of the surveys are important here: their timing, its consistency and the effective forecasting spans involved; changes

This paper is limited to one phase of this large research project, namely the search for evidence on the extent and locus of those errors that appear to by "systematic." What are the frequencies and significance of bias and autocorrelated errors? How do the findings vary for different variables and predictive horizons? For individual and composite forecasts? What do the results indicate about the rationality hypothesis as applied to macroeconomic predictions?

The next section defines the measures to be used, discusses problems with the data and presents the evidence on the question of bias in multiperiod predictions by individuals. Section III addresses the problem of serially dependent residual errors and applies the tests for unbiasedness to group forecasts from the surveys. Section IV deals with the tests for autocorrelation in the "knowable" forecast errors. The final section (V) sums up the results and places them in the context of earlier related work.

II. Testing for Bias in Multiperiod Predictions

The Actual and Predicted Values Defined

Let $t = 1, \ldots, n$ be the survey quarter during which the forecast is made and $t + j$ be the target quarter to which the forecast refers, where $j = 0, \ldots, 4$ quarters. For any variable, $A_{jt} = A_{t+j}$ denotes the actual level in the target period and $P_{ijt} = P_{i,\ t+j}$ denotes the corresponding level prediction by the i^{th} forecaster. Where appropriate, the actual percentage

---

in composition over time; the role of outliers; and reporting errors. A careful proofreading of the survey questionnaire is needed to detect simple mistakes of calculation, copying, and typing which chance or neglect will always occasion in some replies. The voluminous NBER-ASA materials were submitted to such an audit with the aid of the computer and, where needed, inspection of the original submissions. Although the number of the thus identified mistakes turned out to be very small in relative terms, failure to eliminate them would have affected adversely the evaluation of several individual records.

change is

$$(4) \qquad \dot{A}_{jt} = \left(\frac{A_{t+j} - A_{t+j-1}}{A_{t+j-1}}\right)100, \quad j = 0, \ldots, 4,$$

and the predicted percentage change is

$$(5) \qquad \dot{P}_{ijt} = \begin{cases} \left(\frac{P_t - A^*_{t-1}}{A^*_{t-1}}\right)100, & \text{if } j = 0 \\ \left(\frac{P_{t+j} - P_{t+j-1}}{P_{t+j-1}}\right)100, & \text{if } j = 1, \ldots, 4. \end{cases}$$

The ASA-NBER surveys are taken in the first half of each quarter, at a time when the most recent data available would be the preliminary estimates for the preceeding quarter, which are marked A^*_{t-1} in (5).[8] Consequently, the P figures for the current quarter (j = 0) are authentic ex ante forecasts whose span is approximately one quarter.

The "actual" values are not well defined for many economic variables, such as GNP and components, which are subject to several, often sizable, revisions. Here they are represented by the last data available prior to the benchmark revisions of January 1976 and December 1980. These are presumably the "best" of those estimates that are conceptually comparable to the corresponding survey predictions.[9]

---

[8]An exception is the unemployment rate series which is available monthly.

[9]This procedure imposes on the forecasters the burden to predict future revisions that are assumed to remove observational errors. An alternative is to compare the forecasts with provisional data that are closer to the most recent figures that were available to the forecaster. The most informative approach is one that integrates the analysis of data errors and of predictive errors, which would be a good task for another paper. On the role of preliminary data and revisions in economic measurement and prediction, see Cole, 1969; Howrey, 1978; and Zarnowitz, 1979 and 1982a.

As shown by (5), the base of any change forecast for $j = 0$ is the preliminary estimate of the previous level, A^*_{t-1} (itself a prediction or extrapolation based on imcomplete data). For $j > 0$, the base is the forecast of the level in the preceding quarter, P_{t+j-1}. The differences between the successive levels predicted in a multiperiod forecast made at time t, $P_{t+j} - P_{t+j-1}$, are implicit predictions of changes over the successive subperiods covered. Note that each of these marginal ("intraforecast") predictions covers a single quarterly interval, so the target periods do not overlap. The predicted changes refer to successive quarters, 0-1, 1-2, ... (In contrast, forecasts of average changes over increasing spans, 0-1, 0-2, ..., involve overlapping target periods and their errors are therefore necessarily intercorrelated. See Zarnowitz, 1967, pp. 64-70.)

Estimating Regressions of Actual on Predicted Values

Regressions of the actual on the predicted values have been computed for each of the 79 individuals who participated in at least 12 surveys and also for the series of means of the corresponding predictions (called the group mean forecasts). For the unemployment rate (UR) and inventory investment (CBI), levels were used as in

$$(6) \qquad A_{jt} = a_{ij} + b_{ij}P_{ijt} + u_{ijt} \ , \quad j = 0, \ldots, 4 \ , \quad t = 1, \ldots, n \ ,$$

while for nominal and real GNP, the price index, and consumer durables (GNP, RGNP, IPD, and CEDG), percentage changes were used as in

$$(7) \qquad \dot{A}_{jt} = a_{ij} + b_{ij}\dot{P}_{ijt} + u_{ijt} \ , \quad j = 0, \ldots, 4 \ , \quad t = 1, \ldots, n \ .$$

Estimation of either (6) or (7) requires certain assumptions about the probability distribution of the disturbances u_{ijt}. The simplest and most common approach is to assume that $E(u_t) = 0$, $\mathrm{var}(u_t) = \sigma^2$, and

u_1 ..., u_n are independently distributed, for any i^{th} forecaster and j^{th} target quarter. The technique of ordinary least squares (OLS) applies in this case. The sample least-squares estimates a and b (the subscripts may now be dropped for simplicity) lend themselves to statistical tests of the joint null hypotheses that the true (population) parameters of the relation between A and P are $\alpha = 0$ and $\beta = 1$. A sufficiently high F ratio refutes that hypothesis, suggesting that the forecast contains some systematic errors.

However, it is uncertain whether the assumption that the u's are serially uncorrelated is appropriate in the analytical situation before us. Consider multiperiod forecasts issued each quarter for a chain of m quarters ahead: clearly, both the actual and predicted values for the cumulative changes during the overlapping intervals (0-1, ..., 0-m) will show autocorrelations of, at least, first to m^{th} order. But it may be possible to circumvent this particular problem by focusing on marginal changes over nonoverlapping single-quarter intervals instead of the average or cumulative changes, as it is done in this paper. It is also important to note that the individual forecast series contain gaps whenever a respondent missed any of the surveys (recall that the criterion for inclusion is a minimum of twelve responses which need not be consecutive). While such gaps reduce the informational contents of the data available for estimating the regressions (6) and (7), they also reduce the probable autocorrelations in the disturbance terms of these equations. It would clearly be improper to try to replace the missing observations (predictions) by any kind of interpolation, since this would amount to augmenting authentic forecasts with artifacts. (The worst thing to do, given our purpose of forecast assessment, would be to use the available actual data to close the gaps.) Forecasters miss surveys essentially at random because of reporting problems (Zarnowitz, 1982c), which means that simply dropping the observations when forecasts are not available should be a reasonable procedure which will cause a loss of efficiency in the OLS estimates but not bias or inconsistency.

In sum, this simplest approach to testing for unbiasedness in the regression framework is arguably justified by the nature of our data and objectives, besides having the advantage of using the entire set of the more regular forecasts at our disposal. Of course, this does not reduce the need to check on the autocorrelations among the disturbances, which can be caused by various factors, notably shocks and/or measurement errors in the actual values that are unanticipated and persist for more than one unit period. In this connection, it will be instructive to pay particular attention to forecast series that have no gaps such as the series of comprehensive group mean predictions, and to apply to them the techniques of generalized least-squares (GLS) estimation.

Distributions of the Regression and Test Statistics

Table 1 presents the evidence from a very large collection of forecasts, including 790 P_{ij} and 1,560 $\dot{P}_{ij}$ series. To provide a background of descriptive statistics, the OLS estimates of the intercepts and slope coefficients in equations 6 and 7 are summarized in columns 1-4. There is a great deal of dispersion in these figures, reflecting partly differences in the ability of the individuals to produce unbiased forecasts and partly differences in time coverage.

The means of $a_{ij}(\bar{a})$ tend to increase with j, the distance to the target quarter, at least from Q0 through Q3, except for RGNP (column 1). In contrast, the means of $b_{ij}(\bar{b})$ typically decrease (column 3). The standard deviations of a_{ij} and b_{ij} both tend to rise as the predictive horizon lengthens (see columns 2 and 4, and note the main irregularities in the SD_a figures for IPD and the SD_b figures for CBI). Hence the relative dispersion measures for a_{ij} and b_{ij} behave quite differently: the $SD_a/\bar{a}$ show no

Errata for Zarnowitz's Working Paper #1070

TABLE 1

MULTIPERIOD PREDICTIONS FOR SIX AGGREGATE VARIABLES BY
79 PARTICIPANTS IN ASA-NBER SURVEYS, 1968-1979:
DISTRIBUTIONS OF REGRESSION STATISTICS AND TESTS OF BIAS

Quarter Predicted	Mean Values of Individual Statistics[a]					Percent of Forecasts with Significant Tests[b]				
	$\bar{a}$ (1)	SD_a (2)	$\bar{b}$ (3)	SD_b (4)	$\bar{F}$ (5)	F (6)	t_α (7)	t_β (8)	$F(s)$ (9)	$F(L)$ (10)
					GNP in Current Dollars (GNP)					
Q0	.38	.76	.87	.32	1.46	12.7	15.2	12.7	0	21.7
Q1	.81	.75	.65	.34	1.58	10.1	17.7	19.0	3.0	15.2
Q2	1.17	.88	.52	.41	1.84	11.4	26.6	17.7	3.0	17.4
Q3	1.27	1.06	.46	.51	1.82	16.5	26.6	24.1	3.0	26.1
Q4	1.12	1.24	.50	.59	1.58	11.4	20.2	16.5	0	19.6
					Implicit Price Deflator (IPD)					
Q0	.42	.42	.81	.33	2.63	26.6	19.0	17.7	11.8	37.8
Q1	.72	.51	.69	.42	3.68	46.8	36.7	17.7	20.6	66.7
Q2	1.03	.48	.48	.42	4.36	57.0	48.1	20.3	23.5	82.2
Q3	1.20	.46	.36	.41	4.52	64.6	43.0	17.7	20.6	97.8
Q4	1.27	.66	.42	.65	4.39	58.2	38.0	16.5	8.8	95.6
					GNP in Constant Dollars (RGNP)					
Q0	-.12	.36	1.06	.31	1.60	10.1	19.0	12.7	2.9	15.6
Q1	-.29	.48	1.04	.43	1.64	8.9	7.6	8.9	0	15.6
Q2	-.11	.65	.80	.59	1.60	8.9	7.6	8.9	0	15.6
Q3	.01	.72	.62	.64	1.86	12.7	2.5	7.6	0	22.2
Q4	-.27	1.19	.72	1.04	2.20	15.2	0	12.7	0	26.7
					Unemployment Rate (UR)					
Q0	-.01	.26	1.00	.05	1.17	2.5	3.8	3.8	0	4.3
Q1	-.01	.76	1.01	.14	1.01	2.5	2.5	3.8	0	4.3
Q2	.29	1.17	.98	.22	1.18	3.8	3.8	3.8	0	6.5
Q3	1.01	1.38	.88	.26	1.92	12.7	7.6	8.9	3.0	19.6
Q4	1.80	2.12	.75	.39	1.98	10.1	16.5	11.4	0	17.4
					Consumer Expenditures--Durable Goods (CEDG)					
Q0	.99	.87	.93	.45	2.15	20.0	12.0	13.3	6.5	29.5
Q1	1.26	1.16	.43	.55	1.56	6.7	8.0	16.0	0	11.4
Q2	1.55	1.16	.27	.67	1.38	8.0	8.0	10.7	3.2	11.4
Q3	1.41	1.70	.26	.82	1.16	2.7	2.7	13.3	3.2	2.3
Q4	.57	1.88	.59	.92	.92	4.0	5.3	4.0	0	6.8
					Change in Business Inventories (CBI)					
Q0	2.76	3.54	.88	.52	1.77	16.2	20.3	8.9	11.8	19.6
Q1	1.81	4.54	.93	.61	1.62	10.0	13.9	13.9	2.9	15.2
Q2	2.18	5.61	.82	.80	1.28	6.3	10.1	5.1	5.9	6.5
Q3	2.22	5.72	.78	.74	1.15	3.8	6.3	3.8	0	6.5
Q4	2.97	4.84	.78	.59	1.11	3.8	6.3	2.5	2.9	4.3

[a]The entries in columns 1 and 2 are the means ($\bar{a}$) and standard deviations (SD_a) of the a_{ij} estimates from the regressions of actual values on the individual forecasts. The entries in columns 3 and 4 are the means ($\bar{b}$) and standard deviations (SD_b) of the b_{ij} estimates from the same regressions. See text and equations 6 and 7. The regressions are estimated by ordinary least squares. $\bar{F}$ (column 5) denotes the average values of the F ratios for the tests of H_0: $\alpha = 0$ and $\beta = 1$ performed on the series of individual forecasts for each of the categories covered. All figures refer to those individuals who participated in at least 12 surveys: 75 for CEDG, 79 for each of the other variables.

[b]The significance level is 5% for all tests. The percentages in columns 6-8 refer to all participants in at least 12 surveys (same coverage as in columns 1-5); column 9 to those who responded to 12-19 surveys (31-34); and column 10 to those who responded to 20 or more surveys (44-46). The F tests are for the joint null hypothesis that $\alpha = 0$ and $\beta = 1$, the t_α tests for the hypothesis that $\alpha = 0$, and the t_β tests for the hypothesis that $\beta = 1$.

TABLE 1

MULTIPERIOD PREDICTIONS FOR SIX AGGREGATE VARIABLES BY 79 PARTICIPANTS IN ASA-NBER SURVEYS, 1968-1979: TESTS OF H_0: $\alpha = 0$ and $\beta = 1$, ORDINARY LEAST SQUARES (OLS)

Quarter Predicted[a]	No. of Forecasts[b] (1)	Percentage of Forecasts[c] with: F ratios for $\alpha = 0$, $\beta = 1$ Significant at the level of 1% (2)	5% (3)	10% (4)	t ratios for $\alpha = 0$, Significant at the level of 1% (5)	5% (6)	10% (7)	t ratios for $\beta = 1$ Significant at the level of 1% (8)	5% (9)	10% (10)
GNP in Current Dollars (GNP)										
Q0	79	3.8	12.7	19.0	7.6	15.2	21.5	5.1	12.7	17.7
Q1	79	3.8	10.1	16.5	5.1	17.7	22.8	5.1	19.0	21.5
Q2	79	2.5	11.4	19.0	5.1	26.6	36.7	5.1	17.7	35.4
Q3	79	2.5	15.2	27.8	6.3	26.6	34.2	7.6	24.1	32.9
Q4	79	2.5	11.4	16.5	8.9	20.2	22.8	7.6	16.5	24.1
Implicit Price Deflator (IPD)										
Q0	79	6.3	25.3	39.2	5.1	19.0	29.1	3.8	17.7	20.3
Q1	79	13.9	44.3	62.0	7.6	36.7	49.4	3.8	17.7	25.3
Q2	79	20.2	53.2	79.8	15.2	48.1	60.8	5.1	20.3	31.6
Q3	79	26.6	59.5	77.2	15.2	43.0	60.8	5.1	17.7	29.1
Q4	79	22.8	54.4	73.4	12.7	38.0	50.6	5.1	16.5	21.5
GNP in Constant Dollars (RGNP)										
Q0	79	2.5	10.1	20.2	1.3	19.0	29.1	2.5	12.7	22.8
Q1	79	2.5	8.9	17.7	1.3	7.6	20.3	3.8	8.9	13.8
Q2	79	1.3	8.9	13.9	0	2.5	10.1	1.3	8.9	17.7
Q3	79	2.5	11.4	17.7	0	2.5	3.8	5.1	7.6	16.5
Q4	79	1.3	15.2	25.3	0	0	8.9	3.8	12.7	20.3
Unemployment Rate (UR)										
Q0	79	0	5.1	12.7	0	3.8	6.3	0	3.8	10.1
Q1	79	0	2.5	8.9	1.3	2.5	7.6	1.3	3.8	10.1
Q2	79	0	3.8	10.1	1.3	3.8	8.9	1.3	3.8	6.3
Q3	79	1.3	10.1	22.8	2.5	7.6	19.0	2.5	8.9	12.7
Q4	79	1.3	8.9	17.7	2.5	16.5	25.3	2.5	11.4	26.6
Consumer Expenditures--Durable Goods (CEDG)										
Q0	75	5.3	16.0	34.7	2.7	12.0	25.3	1.3	13.3	21.3
Q1	75	4.0	5.3	16.0	0	8.0	20.0	4.0	16.0	28.0
Q2	75	1.3	9.3	12.0	0	8.0	17.3	2.7	10.7	18.7
Q3	75	0	2.7	9.3	0	2.7	9.3	1.3	13.3	18.7
Q4	75	1.3	5.3	6.7	1.3	5.3	6.7	1.3	4.0	9.3

TABLE 1 (concluded)

Quarter Predicted[a]	No. of Forecasts[b]	Percentage of Forecasts[c] with: F ratios for $\alpha = 0$, $\beta = 1$ Significant at the level of 1%	5%	10%	t ratios for $\alpha = 0$, Significant at the level of 1%	5%	10%	t ratios for $\beta = 1$ Significant at the level of 1%	5%	10%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Change in Business Inventories (CBI)										
Q0	79	1.3	15.2	30.4	2.5	20.3	25.3	1.3	8.9	16.5
Q1	79	1.3	10.1	21.5	3.8	13.9	19.0	0	13.9	17.7
Q2	79	1.3	6.3	12.7	2.5	10.1	17.7	0	5.1	15.2
Q3	79	0	5.0	10.1	1.3	6.3	8.9	0	3.8	7.6
Q4	79	0	2.5	11.4	1.3	6.3	10.1	0	2.5	6.3
SUMMARY[d]										
Variable										
GNP	395	3.0	12.2	19.8	6.8	21.3	27.6	6.1	18.0	26.3
IPD	395	18.0	47.3	66.3	11.1	37.0	50.1	4.6	18.0	25.6
RGNP	395	2.0	10.9	19.0	0.5	6.3	14.4	3.3	10.1	18.2
UR	395	0.5	6.1	14.4	1.5	6.8	13.4	1.5	6.3	13.2
CEDG	375	2.4	7.7	15.7	0.8	7.2	15.7	2.1	11.5	19.2
CBI	395	0.8	7.8	17.2	2.3	11.4	16.2	0.3	6.8	12.7

[a] Q0 denotes the current (survey) quarter, Q1 the following (first future) quarter, etc. The number of the surveys covered is 42 for Q0, 41 for Q1, 40 for Q2, 39, for Q3, and 34 for Q4.

[b] Includes those individuals who participated in at least 12 surveys (75 for CEDG, 79 for each of the other variables).

[c] For the number of forecasts on which the percentages in columns 2-10 are based, see the corresponding entry in column 1. For the explanation of the test statistics, see text.

[d] Contains each individual's predictions for five target quarters (Q0-Q4). The number of forecasts is 395 (79 × 5) for each variable except CEDG (where it is 375 = 75 × 5).

common pattern of change, while the $SD_b/\overline{b}$ ratios increase strongly from Q0 to Q4, with few exceptions.

When the F ratios are averaged across comparable regressions for the individuals, the resulting mean values seem low for all but one of the variables covered, ranging from 0.9 to 2.2 and averaging 1.5 with a standard deviation of .35 (column 5). For the IPD inflation forecasts, however, the $\overline{F}$ values average 3.9 and rise from 2.6 in Q0 to 4.5 in Q3.

The impression of a sharp contrast between the predictions of inflation and those of other variables is confirmed by the relative frequencies of the individual forecast series that failed to pass the joint test for unbiasedness ($\alpha = 0$ and $\beta = 1$) according to the F tests at the 5% significance level (column 6).[10] For IPD, about half of the computed F ratios exceed the critical $F_{.95}$ values, whereas for GNP and RGNP the corresponding frequencies are 12 and 11 percent, for UR, CEDG, and CBI six to eight percent.

According to the separate t tests for regression intercepts and slopes, which also use the significance level of 5%, the incidence of $\alpha \neq 0$ is much higher for IPD than for GNP, while the incidence of $\beta \neq 1$ is similar for the two variables (columns 7 and 8).[11] These tests suggest that the poor overall results for the inflation forecasts, as evidenced by the F ratios, are associated to a larger extent with the deviations of α from zero than with the deviations of β from unity. The t_α tests are also relatively unfavorable to the inventory investment (CBI) forecasts, but for the real

---

[10] In each of these joint tests on two regression coefficients, if the null hypothesis is true, the test statistic should have an F distribution with two degrees of freedom in the numerator and $n-2$ in the denominator (where n, the number of observations varies across the individuals).

[11] The appropriate tests are two-tailed. If the null hypothesis holds, the test statistic should follow the t distribution with $n-2$ degrees of freedom.

growth and consumer durables (RGNP and CEDG) forecasts it is the results of the t_β tests that appear to be more damaging.

The test results do not show a common pattern of systematic dependence on the time horizon j. Thus for IPD the frequencies of significant F and t_α ratios increase sharply between Q0 and Q2 or Q3, but those of the t_β ratios do not (columns 6-8). The frequencies for UR generally tend to rise, those for CEDG and, particularly, CBI tend to decline as the target quarter recedes into the future. The figures for the other variables show on the whole smaller or more irregular fluctuations.

The Effects of Sample Size

Although broadly based and rich in comparison with the few small samples used in most studies of economic forecasts, our data also have some important limitations that need to be recognized. The forecast series are numerous but inevitably much shorter than would be desirable, since our surveys began in 1968 only. The minimum requirement of participation in at least twelve surveys improves the data by eliminating the occasional respondents and the shortest series.[12] As a result, the distributions of the admitted forecast sets are skewed toward the longer series. But the average number of observations per series is still no more than 23, with a standard deviation of 8.

The conventional 1% and 5% significance levels imply low (.01 and .05) probabilities of wrongly rejecting the null hypothesis H_0 when it is true but also high complementary (.99 and .95) probabilities of wrongly accepting H_0 when it is false. For small sample sizes, therefore, these tests have very low

---

[12] There are a few exceptions where a series contains less than twelve observations. These refer to the longer horizons and arise because some forecasters occasionally predicted fewer than four quarters ahead. Thus of the 395 $\dot{P}_{ij}$ series for GNP, 16 (4%) have 10 or 11 observations each, all but four of them for Q4. (The count is very similar for each of the other variables.)

power against the alternative composite hypothesis which is merely a negation of H_0 (i.e., H_1: $\alpha \neq 0$, $\beta \neq 1$). This raises a serious question about the meaning of the test results in such cases.[13]

A simple experiment strikingly illustrates the importance of the sample size in this context. The frequencies of the F ratios that are significant at the 5% level are throughout very much lower for the forecasters who participated in 12 to 19 surveys than for those who participated in 20 or more surveys (Table 1, columns 9 and 10). Indeed, the proportions for the first subset, $F(s)$, are typically zero or less than five percent and average 1.9, except for IPD where they range between 9 and 24 percent and average 17.1. In contrast, the proportions for the second subset $F(\ell)$, are concentrated between 10 and 25 percent and average 14.4, except again for IPD where they range between 38 and 98 percent, and average 76.0! Clearly, had only the shorter series been at our disposal, they would have led us to an overly favorable appraisal of the forecasts, though not without a correct warning about the relatively high incidence of bias in the predictions of inflation. It should be noted that the predictions of both groups of forecasters, those with the shorter (s) and those with the longer (ℓ) series, are spread about equally across the 1968-79 period, so that the large discrepancies between the reported results for $F(s)$ and $F(\ell)$ cannot be attributed to differences in the periods covered.[14]

---

[13]As shown in Zellner, 1979, several issues arise in analyzing regression hypotheses, notably the asymmetric treatment of H_0 and H_1 in classical tests, the associated uncertainty about the choice of significance levels that are appropriate for different sample sizes, and the "sharpness" of null hypotheses. Although the problems are well known in principle, they are seldom given much attention in textbooks and are almost habitually disregarded in applied economic and econometric literature.

[14]The shorter series number 31-33, the longer series 44-46, depending on the variable covered (see Table 1, note b for more detail). For the 42 surveys of 1968:4-1979:1, the mean (standard deviation) of the participation numbers is 43 (9); for the two subsets of 21 surveys each, 1968:4-1973:4 and 1974:1-1979:1,

To increase power, higher significance levels may be employed. Table 2 shows that the F(s) frequencies at the 10% level exceed their counterparts at the 5% level by factors ranging from 3 to 14. In contrast, the F(s) frequencies at the 1% level are all zero, misleadingly suggesting that no bias at all exists in this group of relatively short forecast series (cf. columns 2, 5, and 8). For the longer series, however, the decision to use 10% instead of 5% as the significances level would have made little difference in our conclusions, and even at the 1% level the negative results on the inflation forecasts are very evident in the $F(\ell)$ entries (columns 3, 6, and 9). For the total sample, too, the high incidence of bias in the IPD predictions stands out everywhere, but here the comparisons are much less favorable to the other variables at the 10% than at the lower significance levels (columns 1, 4, and 7).

Confidence Regions

Consider the ratio

$$(8)\quad F = \frac{1}{2(c_{11}c_{22} - c_{12}^2)s_u^2}\left[c_{11}(b - \beta)^2 + c_{22}(a - \alpha)^2 + 2c_{12}(a - \alpha)(b - \beta)\right] ,$$

where s_u^2 is the variance of the calculated regression residuals and c_{ij} is the (i, j)th element in the variance-covariance matrix of the estimated coefficients, divided by s_u^2. The confidence region for α and β is given for any selected confidence coefficient g (say, .95) by $F \leqslant F_g$, where the probability $P(F < F_g) = g$. It is an ellipse centered at (a, b), and the higher g the larger is the ellipse. In the present context, it is of interest to compare the confidence regions for selected "short" and "long" series of

---

the corresponding figures are 48 (8) and 38 (8), respectively. Thus some attrition occurred in the number of forecasters per survey. However, its effect was about the same for the two groups of forecasters: for set s, the proportion of observations in the earlier period is 0.61, for set l, it is 0.64.

TABLE 2
SUMMARY OF RESULTS FOR TESTS OF H_0: $\alpha = 0$, $\beta = 1$,
TWO GROUPS OF FORECASTERS, SIX VARIABLES, 1968-1979

	Percent of Forecasts with F ratios That are Significant								
	At the 1% Level			At the 5% Level			At the 10% Level		
Variable	F	$F(s)$	$F(\ell)$	F	$F(s)$	$F(\ell)$	F	$F(s)$	$F(\ell)$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
GNP	3.0	0	5.2	12.4	1.8	20.0	21.0	11.5	27.8
IPD	19.2	0	33.8	50.6	17.1	76.0	69.1	46.5	86.2
RGNP	2.3	0	4.0	11.1	0.6	19.1	20.5	8.2	29.8
UR	0.5	0	0.9	6.3	0.6	10.4	15.4	8.5	20.4
CEDG	2.3	0	4.1	8.3	2.6	12.3	14.7	9.0	18.6
CBI	0.8	0	1.3	8.0	4.7	10.4	17.2	14.7	19.1

NOTE: The symbols for the variables are identified in Table 1. The entries in columns 1, 4, and 7 refer to all individuals who participated in at least 12 of the quarterly ASA-NBER surveys in the 1968:4-1979:1 period (75 for CEDG, 79 for each of the other variables). The entries in columns 2, 5, and 8 refer to those who responded to at least 12 but fewer than 20 of the surveys (31 for CEDG, 34 for IPD and RGNP, and 33 for each of the other variables). The entries in columns 3, 6, and 9 refer to those who responded to 20 or more of the surveys (44 for CEDG, 45 for IPD and RGNP, and 46 for each of the other variables).

forecasts from our collection and observe how they vary with the choice of g and relative to the (0, 1) point of the null hypothesis.

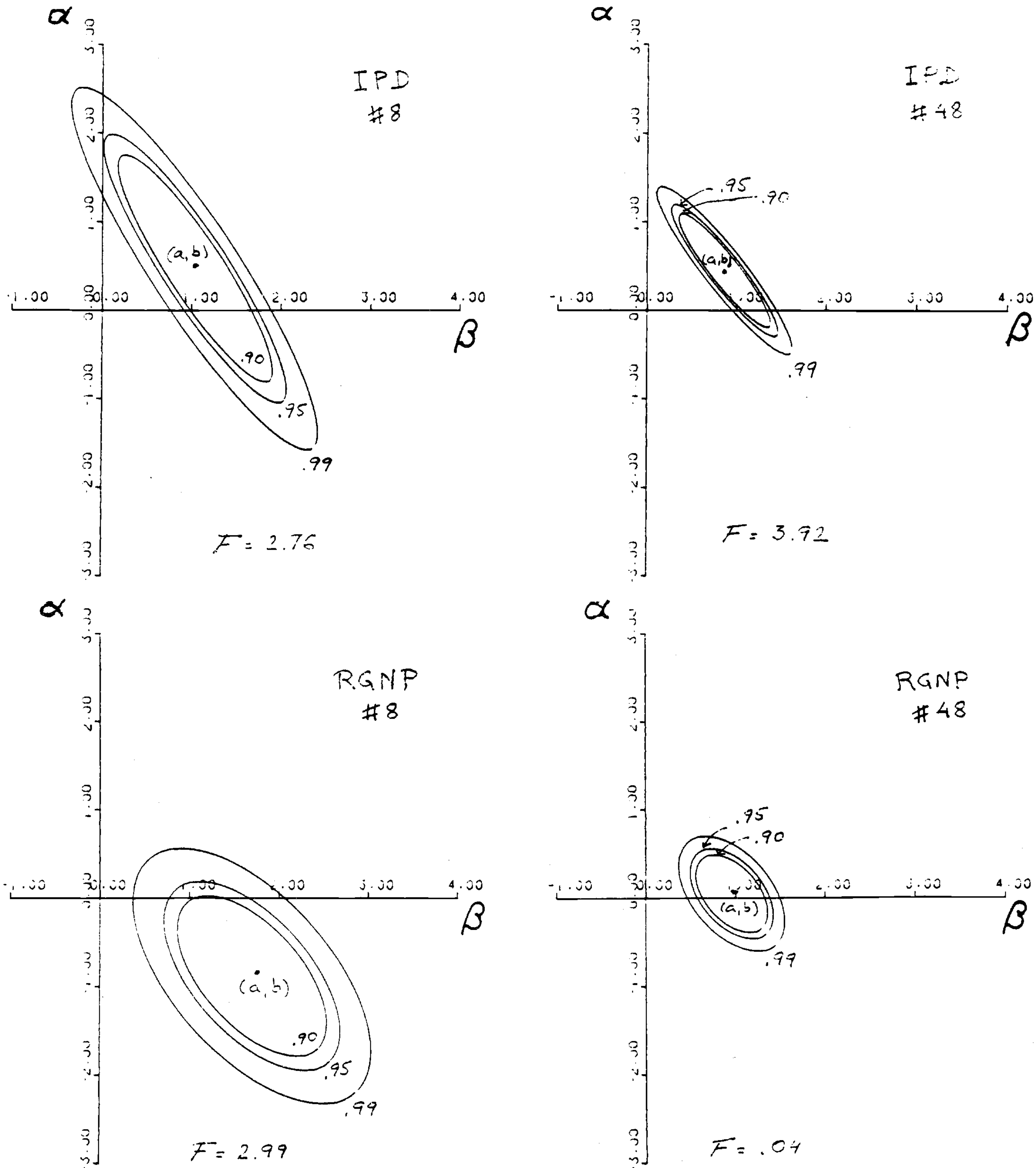
For purposes of illustration, two forecasters were chosen, one coded "8" who participated in 13 consecutive surveys, 1972:1-1975:1, the other "48" whose record includes 33 consecutive surveys, 1968:4-1976:4. Using their Q0 forecasts of inflation and real growth, Chart 1 demonstrates the strong dependence of the results on the sample size. For either variable, the ellipses for the shorter series are much larger than those for the longer series (about twice as long and twice as wide as measured by the major and minor axes). Had space been available for more such comparisons, they would generally confirm the large gains in the precision of numerical statements that can thus be derived for the longer forecast series.

The concentric ellipses associated with the confidence coefficients of .99, .95, and .90 (which correspond to the significance levels of 1%, 5%, and 10% in our tests of H_0: $\alpha = 0$, $\beta = 1$) are close to each other for the longer series, spaced more widely apart for the short ones. Although understandably motivated by the wish to reduce the probability of type I errors, the use of high g values in analyzing small sets of predictions can be quite costly in terms of the lack of precision implied by large confidence regions.

The high incidence of bias in the inflation forecasts is on the whole reaffirmed by this analysis, as exemplified by the IPD graphs in Chart 1. Here the (0, 1) points are located very near the boundaries of the confidence regions for both forecasters: within the ellipses for the 1% level of significance but barely inside or outside those for the 5% and 10% levels. In contrast, (0, 1) is near the center of the ellipses for the RGNP growth rate forecasts in the

CHART 1

Confidence Regions for Selected Forecasts of
Inflation (IPD) and Real Growth (RGNP)



case of the long series #48 but on the periphery or outside in the case of the short series #8.[15]

That the confidence ellipses in Chart 1 have downward sloping major axes indicates that a and b are negatively correlated, which simply reflects the fact that the mean values of the forecasts are positive.[16]

Mean Errors

The tests summarized in Tables 1 and 2 suggest the presence of certain systematic errors in some of the forecasts. An analysis of the distributions of the mean errors of the forecasts helps to identify the probable nature of such errors.

A tendency toward underestimation of change has long been observed in a great variety of forecasts; it is consistent with rational expectations, but it also can arise in biased predictions. Table 3 shows that almost all forecasters underestimated inflation, and did so increasingly for the more distant future. In contrast, real growth as measured by the rates of change in RGNP was predominantly overestimated in this period of an unexpected deterioration in both inflation and the cyclical business performance. On the average, these overestimates rise steadily with the predictive horizon. The underestimates of the price component and the overestimates of the quantity component tend to cancel each other in the predictions of rates of change in current-dollar GNP, where the means errors are negative for most individuals but on the average

---

[15]The critical values $F_{.99}$, $F_{.95}$, and $F_{.90}$ are 2.86, 3.98, and 7.24, respectively, for the smaller sample; the corresponding values for the larger sample are 2.48, 3.31, and 5.36. The calculated values of F are listed on Chart 1.

[16]An elementary property of the two-variable regression model is that cov(a, b) = $-\bar{x}$ var b, where $\bar{x}$ is the mean of the explanatory variable. In our regressions the forecasts play the role of x.

TABLE 3

SELECTED STATISTICS ON THE DISTRIBUTION OF MEAN ERRORS IN INDIVIDUAL FORECASTS, 1968-1979

Quarter Predicted	GNP (1)	IPD (2)	RGNP (3)	UR (4)	CEDG (5)	CBI (6)
			Means (Standard Deviations) of the Mean Errors[a]			
Q0	-.12(.21)	-.16(.14)	.04(.24)	.04(.05)	-.92(.67)	-2.40(1.95)
Q1	-.07(.20)	-.30(.17)	.23(.22)	-.01(.11)	-.36(.70)	-1.88(2.28)
Q2	-.13(.19)	-.39(.18)	.26(.23)	-.12(.17)	-.27(.77)	-1.39(2.82)
Q3	-.13(.21)	-.49(.17)	.35(.25)	-.29(.23)	.03(.76)	-1.10(3.10)
Q4	-.08(.29)	-.61(.21)	.53(.31)	-.32(.27)	.14(.80)	-1.85(2.80)
			Percentage of Under (Over) Estimates[b]			
Q0	71 (29)	89 (11)	34 (66)	14 (86)	91 (9)	95 (5)
Q1	63 (37)	96 (4)	11 (89)	47 (53)	64 (36)	85 (15)
Q2	76 (24)	98 (2)	14 (86)	80 (20)	65 (35)	71 (29)
Q3	73 (27)	99 (1)	10 (90)	92 (8)	52 (48)	69 (31)
Q4	62 (38)	99 (1)	2 (98)	86 (14)	41 (59)	73 (27)

Variables Predicted: GNP, IPD, RGNP, UR, CEDG, CBI.

[a]The errors are defined as predicted minus actual value, so minus (plus) signs are associated with under (over) estimates. For GNP, IPD, and CEDG, the mean error is computed in percentage change terms as $\dot{P}_{ij} - \dot{A}_{ij}$; for UR and CBI, it is computed in terms of levels as $\bar{P}_{ij} - \bar{A}_{ij}$, for any i^{th} individual and j^{th} target quarter. (See text and equations 4 and 5 above for definitions of P_{ijt}, A_{jt}, $\dot{P}_{ijt}$, and $\dot{A}_{jt}$; the bars indicate averaging over time t.) The means of the mean errors across the individuals are without, the corresponding standard deviations are within the parentheses.

[b]The percentage of individual forecasters with mean errors that are negative (positive) is shown without (within) the parentheses. The number of individuals covered is 75 for CEDG, 79 for each of the other variables (all forecasters who participated in at least 12 quarterly ASA-NBER surveys in the period 1968:4-1979:1).

very small throughout (cf. columns 1-3). Underpredictions prevail for the unemployment rate in Q2-Q4 (consistent with the overprediction of real growth) and for business inventory investment, while the record for the rates of change in consumer durables is more mixed (columns 4-6).

III. Allowing for Serially Dependent Residual Errors

Autocorrelated Disturbances and Bias in Individual Forecasts

Tests for serial correlations among the regression residuals u_{ijt} (eqs. 6 and 7) have been made for all those series in our collection that consist of at least 13 observations and contain no gaps. These data refer to the forecasts by 18-20 individuals (the number varies somewhat depending on the target) who participated in more than 12 consecutive surveys. The nonconsecutive predictions by the same forecasters are omitted. The series number 452, vary in length from 13 to 33 and average 19 quarters, and cover Q0-Q3 (the samples for Q4, which are smaller, are not included).

For each of the thus obtained residual error (u_t) series, serial correlation coefficients $\hat{\rho}_k$ are computed for $k = 1, \ldots, 6$. (Since many of the available series are short, only the first six coefficients are considered.) On the assumption of homoscedasticity, these measures are defined as

$$\hat{\rho}_k = \mathrm{cov}(u_t, u_{t-k})/\mathrm{var}(u_t) \quad (9)$$

The Box-Pierce statistic Q serves as a convenient test for the presence of autocorrelation in such sets of the $\hat{\rho}$'s . In the present context, it is calculated by

(10) $$Q = n(n + 2) \sum_k^6 (n - k)^{-1} \hat{\rho}_k^2 ,$$

which is approximately distributed as chi-square with six degrees of freedom.[17]

Most of the Q statistics computed for the inflation and unemployment forecasts are found to be statistically significant at the 5% and 10% levels, and the frequencies are particularly high for IPD (see Table 4, columns 1-4). In contrast, only about one-sixth of the F tests for RGNP produces similar results, and the frequencies for CEDG are not much higher. According to these figures, then, the incidence of autocorrelated residual errors varies greatly across the variables covered.[18]

We next match up for each individual the results of the Q tests with those of the previously discussed F tests and show the percentage distribution of the forecasts according to the significance (at the 10% level) of both statistics (Table 4, columns 5-8). Because the F tests are based on larger samples that include nonconsecutive observations for the same forecasters, the measures underlying this cross-tabulation are not strictly comparable, but the broad indications obtained are deemed to be meaningful and of sufficient interest.[19]

---

[17]If the errors formed random uncorrelated sequences, the $\{\hat{\rho}_k\}$ would themselves be uncorrelated with variances equal to $(n - k)/n(n + 2)$. For large values of n and relatively small m, the variances approximate $1/n$ and $Q = n \sum_k^m \hat{\rho}_k^2 \sim \chi_m^2$. In view of the small size of the available samples, it seemed advisable to avoid these common approximations. See Box and Pierce, 1970.

[18]The frequencies of significant Q's increase from Q0 to Q3 for IPD, UR, and CBI, but appear to be unrelated to the predictive horizon for the other variables.

[19]Given the nature of the available data, few alternatives to the adopted procedures were perceived and none seemed preferable in terms of the prospective costs and returns.

TABLE 4

FREQUENCIES OF SIGNIFICANT Q AND F STATISTICS
FOR SELECTED FORECASTS, SIX VARIABLES, 1968-1979

Variable	No. of Forecast Series[a]	Signif. Level of Statistics[b]			Significance at the 10% Level[c]			
		1%	5%	10%	Q and F	Q only	F only	Neither
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Percent of Forecasts[d]						
GNP	75	12.0	21.3	35.3	6.6	25.0	22.4	46.0
IPD	80	47.5	68.8	73.8	66.2	7.5	17.5	8.8
RGNP	75	5.3	16.0	17.3	4.0	13.3	21.3	61.3
UR	71	33.8	53.5	62.0	14.1	47.9	0	38.0
CEDG	71	2.8	11.3	22.5	5.0	37.5	7.5	50.0
CBI	80	17.5	32.5	43.8	8.5	14.1	18.3	59.2

[a]Each series includes consecutive observations only, by participants in more than 12 surveys.

[b]Refers to the Box-Pierce statistics as defined in eq. (10) with 6 degrees of freedom. See text.

[c]Refers to the set of Q and F statistics matched by individuals, as explained in the text. Except for rounding, the sum of the corresponding entries in columns 5-8 is 100.0.

[d]Based on entries in column 1.

Serial correlation in the error terms u_t may bias upward the F statistics, causing them wrongly to reject the null hypothesis. But cases in which both Q and F are significant represent only four to eight percent of our observations for GNP, RGNP, CEDG, and CBI, and 14 percent for UR (column 5). Once more, the situation is entirely different for IPD, where such cases account for as much as 66 percent of the forecasts. Except for IPD and UR, the F's clearly are more likely to be significant when the F's are not (cf. columns 5 and 7). Often, too, the Q's are significant while the F's are not; this is so in particular for GNP, UR, and CEDG (column 6). Finally, except for IPD, tests which find neither Q nor F to be significant are very frequent, adding up to more than half of the observations (column 8).

OLS Estimates and Tests for the Group Mean Forecasts

Consider now the overall group forecasts, that is, series of means of the corresponding predictions by all individuals included in this study. For each of our thirty target categories (6 variables × 5 horizons), actual values are regressed on these composite forecasts by means of ordinary least squares. Table 5 shows that the results vary greatly for the different targets. The absolute values of the regression intercepts $|a|$ often increase with the predictive horizon, while the signs of these estimates are about equally mixed (column 1). All of the slope coefficients (b) are positive but they otherwise display no common regularities (column 2). For example, the b's tend to be smaller than 1.0 and declining with the horizon for IPD and UR, larger than 1.0 and rising with the horizon for RGNP and CBI.

For GNP, the values of a do not deviate significantly from zero and the values of b from unity, according to the F and t ratios (columns 3-5). In contrast, the F tests strongly reject H_0: $(\alpha, \beta) = (0, 1)$ for the inflation (IPD) forecasts, particularly in the more distant quarters, and the t

TABLE 5

REGRESSIONS OF ACTUAL ON PREDICTED VALUES AND TESTS OF BIAS,
30 SERIES OF GROUP MEAN FORECASTS, OLS ESTIMATES, 1968-1979

Quarter Predicted	Regression Estimates[a] a (1)	b (2)	*F* ratio[b] for $\alpha = 0, \beta = 1$ (3)	t tests[c] for $\alpha = 0$ (4)	$\beta = 1$ (5)	Squared Correlation[d] $\bar{r}^2$ (6)	Durbin-Watson[e] DW (7)	Additional Statistics[f] St. Error SER (8)	Dependent Variable Mean (9)	S.D. (10)
				Gross National Product (GNP)						
Q0	-.42	1.26	2.17	-1.28	1.72§	.63	2.36°	.64	2.23	1.06
Q1	-.03	1.06	.28	-.05	.25	.32	2.00°	.87	2.23	1.06
Q2	.31	.91	.39	.49	-.30	.18	1.92°	.96	2.25	1.06
Q3	.23	.95	.36	.29	-.13	.16	1.95°	.99	2.27	1.06
Q4	.02	1.03	.16	.22	.07	.12	1.61°	.95	2.27	1.01
				Implicit Price Deflator (IPD)						
Q0	-.08	1.17	4.37#	-.39	1.26	.64	1.35+	.39	1.53	.64
Q1	.37	.93	5.95*	1.37	-.35	.33	1.06+	.53	1.56	.65
Q2	.75	.67	8.02*	2.41#	-1.35	.14	.79+	.60	1.58	.65
Q3	.94	.56	10.23*	2.81*	-1.66	.08	.69+	.63	1.60	.65
Q4	1.08	.50	11.54*	2.65#	-1.47	.03	.62+	.67	1.67	.68
				Real Gross National Product (RGNP)						
Q0	-.36	1.44	6.02*	-2.70#	3.45*	.75	2.47x	.63	.69	1.26
Q1	-.64	1.54	3.30#	-2.56#	2.22#	.49	2.09°	.91	.66	1.28
Q2	-.70	1.54	1.82	-1.82§	1.38	.26	1.80°	1.10	.66	1.28
Q3	-.58	1.33	1.27	-1.08	.61	.11	1.63°	1.20	.66	1.28
Q4	-1.21	1.80	3.22§	-1.83§	1.26	.16	1.26+	1.13	.59	1.24
				Unemployment Rate (UR)						
Q0	-.05	1.00	2.89§	-.48	.22	.99	1.42+	.15	5.92	1.45
Q1	.05	.99	.01	.17	-.16	.91	.86+	.42	5.97	1.40
Q2	.51	.93	.74	1.04	-.90	.76	.60+	.42	5.97	1.40
Q3	1.35	.81	3.06§	2.13§	-1.84§	.59	.42+	.82	6.09	1.27
Q4	1.98	.71	3.57#	2.45#	-2.22#	.43	.30+	.87	6.22	1.15
				Consumer Expenditures-Durable Goods (CEDG)						
Q0	.08	1.62	8.19*	.17	3.22*	.63	2.18°	2.47	2.30	4.05
Q1	.02	1.25	.47	.02	.56	.14	2.52x	3.81	2.22	4.11
Q2	.62	.84	.16	.47	-.26	.02	2.42°	4.07	2.19	4.11
Q3	.58	.83	.10	.38	-.24	.01	2.45°	4.09	2.20	4.10
Q4	-2.00	2.05	.79	-1.17	1.26	.12	2.30°	3.53	1.96	3.77

Table 5 concluded

Quarter Predicted	Regression Estimates[a] a (1)	b (2)	F ratio[b] for $\alpha = 0, \beta = 1$ (3)	t tests[c] for $\alpha = 0$ (4)	$\beta - 1$ (5)	Squared Correlation[d] $\bar{r}^2$ (6)	Durbin-Watson[e] DW (7)	Additional Statistics[f] St. Error SER (8)	Dependent Variable Mean (9)	S.D. (10)
				Change in Business Inventories (CBI)						
Q0	-.50	1.39	4.05#	-.27	1.99§	.55	1.88°	7.54	9.78	11.23
Q1	-3.15	1.65	4.41#	-1.25	2.47#	.48	1.34+	8.48	10.38	11.80
Q2	-3.98	1.69	3.13§	-1.25	2.09§	.38	1.40+	9.30	10.56	11.81
Q3	-5.94	1.84	3.17§	-1.65	2.27#	.37	1.22+	9.41	10.50	11.82
Q4	-6.21	1.83	2.41	-1.39	1.92§	.32	1.13+	10.07	11.35	12.20

---

[a]The series of levels are used for UR and CBI, series of percentage changes for the other variables. All measures refer to the means of predictions by those respondents to the quarterly ASA-NBER surveys, 1968:4-1979:1, who participated in at least 12 surveys. The series used in each of the regressions for Q0-Q3 contain 42 observations for as many surveys; the series for Q4 contain 37 observations each. See text for further explanations.

[b]The critical values of F (2, 40) at the indicated significance levels are: 6.04 ($\frac{1}{2}$ of 1%), 5.18 (1%), 4.06 ($2\frac{1}{2}$ %), 3.23 (5%), and 2.41 (10%). The critical values of F (2, 35) are slightly higher. * means significant at the 1% level, # at the 5% level, and § at the 10% level.

[c]For the meaning of symbols *, #, and §, see note b above.

[d]Corrected for degrees of freedom.

[e]Corrected for the gaps in the data for Q4. ° means null hypothesis of no first-order serial correlation in the u's is accepted at the 5% significance level; x result indeterminate; + positive serial correlation present. The 5% significance points of d_l and d_u for $n = 40$ and $k = 1$ are: 1.44 and 1.54.

[f]SER: standard error of the regression $(\sum u^2/n - 2)$, SD: standard deviation of the dependent variable (actual values).

statistics suggest that this is attributable mainly to $\alpha > 0$. The estimates for UR show a striking dependence on the horizon but bias is here strongly indicated in the longest forecasts only. Elsewhere, on the contrary, it is the short predictions (Q0 and Q1 for RGNP and CBI, Q0 for CEDG) that are apparently biased, which could be due to measurement errors in estimating the base of the forecast. Here the t ratios often suggest inefficiency in the sense of $\beta > 1$.

As background information, Table 5 includes statistics on the goodness of fit ($\bar{r}^2$), the dispersion of the errors associated with the regression line (SER), and the means and standard deviations of the series of realizations (columns 6, 8-10). These measures are more relevant in evaluating aspects of accuracy rather than rationality of the forecasts, and some of them are treated elsewhere.[20] But it is interesting to observe that the incidence of bias does not appear to be systematically related to either the relative accuracy of the forecasts or the relative smoothness of the target series. Thus, the percentage changes in GNP are far more volatile than the levels of the unemployment rate, which helps to explain why the $\bar{r}^2$ coefficients are so much higher for the latter (compare the corresponding entries in columns 6, 9, 10), but the F and t tests are much more favorable to GNP than to UR. There are strong indications of bias in the forecasts of IPD inflation and none in those of the rates of change in CEDG beyond Q0, but the relative variability of the former series is much less than that of the latter. In general, bias does not imply particularly large errors, and some of the forecasts that appear to be highly

---

[20]See Zarnowitz, 1982c. (Note that there is a small difference in coverage between the two papers, which however has little effect on the results. In the other, earlier, paper the series end in 1979:1 so that the number of observations for Q0-Q3 is 42 - j, j = 0, ..., 3. Here the series are extended so that the number of observations for Q0-Q3 is 42 in each case.)

biased are indeed relatively accurate (notably for UR but also the short predictions of IPD, RGNP, and CBI).

The mean square error of a series of forecasts (say, any of the overall group mean forecasts P_g) can be viewed as a sum of the mean component, slope component, and residual variance defined as

$$(11) \qquad M_p = MC + SC + RV = \bar{e}_p^2 + (1 - b)^2 S_p^2 + S_u^2 ,$$

where $\bar{e}_p$ is the mean error P_g, and S_p and S_u are standard deviations of P_g and of the residual disturbances u from the regressions of A on P_g, respectively.[21] The average proportions of the three components, in percent of the corresponding mean square errors, are tabulated below:

	GNP	IPD	RGNP	UR	CEDG	CBI
$(MC/M_p)100$	2	31	6	5	3	4
$(SC/M_p)100$	2	2	8	3	5	12
$(RV/M_p)100$	96	67	86	92	92	84

Reflecting the favorable results of the bias tests, RV accounts for more than 90% of M_p for GNP, UR, and CEDG. The MC estimates are 6% or less, with the important exception of IPD inflation, where they rise from 15 to 45% between Q0 and Q4. The SC proportions are relatively high for the shortest predictions of RGNP and CBI which are very sensitive to errors in the jump-off estimates; elsewhere they average 2-5% only.

The Durbin-Watson (DW) statistics listed in Table 5, column 7, suggest that the residual disturbances from the regressions of actual on predicted

[21] See Theil, 1965, p. 38, and Mincer and Zarnowitz, 1969, pp. 10-11. Equation 11 applies to any of the regressions and subscripts for variable and horizon are not needed here. The distinction between level and percentage change series is also disregarded to simplify notation.

values for GNP, RGNP, and CEDG are essentially free of first-order serial correlations when 5% significance points are used. On the other hand, the DW tests for IPD and UR indicate strongly the presence of positively autocorrelated residuals, and most of the results for CBI point with less force in the same direction.

The well-known property of positively autocorrelated residuals is to bias downward the SER and upward the $\bar{r}^2$ values (while leaving the OLS regression estimators unbiased and consistent). The loss of efficiency--underestimation of sampling variances of the regression coefficients--may in some cases invalidate the results of our tests, which motivates the next step in this analysis.

Autoregressive Errors and GLS Estimates

Table 6 presents estimates of the parameters in linear regression models with autoregressive errors of the general form

(12)

$$A_t = a + bP_g t + u_t ,$$

$$u_t = \varepsilon_t - \sum_{i=1}^{j} \rho_i u_{t-1} ,$$

where ε_t is a normally and independently distributed error term and j equals 1, 2, and 3 for Q1, Q2, and Q3, respectively.[22]

---

[22]The procedure used is AUTOREG, see SAS/ETS User's Guide, 1980 edition, pp. 8.1-8.7. AUTOREG first estimates the OLS regressions, computes the autocorrelations of the resulting residuals, and uses the Yule-Walker equations to estimate the ρ's. Then the variables from the original data are transformed by the autoregressive model and new estimates of the regression parameters (here a and b) are obtained by an OLS regression using the transformed data. The procedure thus yields generalized least squares (GLS) estimates. It is not applicable to data with missing values, hence the exclusion of Q4 from Table 6.

TABLE 6

REGRESSIONS OF ACTUAL ON PREDICTED VALUES AND TESTS OF BIAS,
18 SERIES OF GROUP MEAN FORECASTS, GLS ESTIMATES, 1968-1979

Quarter Predicted	Regression Estimates a (1)	Regression Estimates b (2)	*F* ratio for $\alpha = 0, \beta = 1$ (3)	*t* tests for $\alpha = 0$ (4)	*t* tests for $\beta = 1$ (5)	Est. Autoregression Coefficients $\hat{\rho}_1$ (6)	$\hat{\rho}_2$ (7)	$\hat{\rho}_3$ (8)	Squared Correlation $\bar{r}^2$ (9)	Durbin-Watson DW (10)
Gross National Product (GNP)										
Q1	-.02	1.06	-.04	.24	.01				.31	1.99°
Q2	.27	.93	.45	.47	-.27	-.02	.17		.19	1.97°
Q3	.15	.99	.44	.21	.03	-.02	.18	.01	.12	1.95°
Implicit Price Deflator (IPD)										
Q1	.60	.75	2.56§	1.49	-.84	-.47*			.10	2.06°
Q2	1.15	.36	2.95§	2.14#	-1.58	-.56*	-.07		.06	2.13°
Q3	1.03	.49	2.27	1.64	-1.04	-.59	-.12	.03	.08	2.12°
Real Gross National Product (RGNP)										
Q1	-.64	1.54	3.58#	-2.67#	2.32#	.06			.50	2.02°
Q2	-.74	1.59	1.98	-1.91§	1.48	-.08	.11		.24	1.92°
Q3	-.28	1.01	.65	-.44	.01	-.18	.03	-.07	.04	1.99°
Unemployment Rate (UR)										
Q1	.41	.93	.46	.88	-.94	-.54*			.78	1.40+
Q2	.85	.87	1.14	1.46	-1.36	-1.03*	.55*		.66	.91+
Q3	2.65	.58	4.99#	3.15#	3.00*	-1.31*	.80*	-.18	.25	.74+
Consumer Expenditures-Durable Goods (CEDG)										
Q1	-.06	1.30	.87	-.08	.80	.29§			.20	2.14°
Q2	-.17	1.25	.43	-.17	.55	.27§	.19		.10	2.09°
Q3	-.68	1.47	.77	-.66	.99	.30§	.21	.19	.12	2.11°
Change in Business Inventories (CBI)										
Q1	-2.56	1.59	2.31	-.77	1.74	-.31#			.33	1.70°
Q2	-3.25	1.60	1.85	-.88	1.56	-.35#	.18		.26	1.87°
Q3	-6.10	1.86	2.57§	-1.50	2.05§	-.44#	.16	.13	.28	1.81°

NOTE: See text and note 22 on the model and computer procedure used. On the meaning of the measures and symbols, see notes to Table 5.

For GNP, none of the estimates of the autoregressive parameters ρ_i are significant, confirming the absence of serial correlation among the residuals from the OLS regressions. Not surprisingly, then, all the statistics in Table 6, lines 1-3, resemble closely their counterparts in Table 5, lines 2-4.[23] The GLS and OLS estimates also show no significant differences for the forecasts of RGNP in Q2 and and Q3 and those of CEDG in Q1-Q3 (all cases in which there is no clear evidence of serially correlated u's).

There is no doubt about the presence of first-order autocorrelations in the error terms of the OLS regressions for inflation and inventory investment, and here the GLS estimation results in large reductions of the test statistics. The F ratios for IPD in Table 6 are much smaller than their counterparts in Table 5 but still significant at the 10% level.

Finally, there is no visible improvement in the cases of RGNP-Q1 and UR-Q3, where the F and t ratios in Table 6 are indeed larger than the corresponding entries in Table 5. It should be noted that the high values of $\hat{\rho}_1$ and $\hat{\rho}_2$ indicate the presence of a second-order autoregressive process in the error terms of the OLS regressions for the unemployment rate in Q2 and Q3.

IV. Testing for Autocorrelation in Forecast Errors

Framework of Analysis and Results for Individual Forecasts

The actual values employed in the previous section include all the nonconceptual (prebenchmark) revisions in the data. These revisions presumably bring the data closer to the "true" values that one would like to have predicted. But it is important to recognize that such data, and hence the estimates

---

[23] Output from the initial OLS part of AUTOREG is identical with the output of the TSP program that was used to generate the corresponding estimates in Table 5 at least up to four decimal places.

derived from them are all ex post in nature. The residual errors from our regressions could not have been known to the forecasters on the current basis. The requirement that such errors be free of serial correlation is therefore not a straightforward test of rationality in the sense of efficient use of contemporaneous information.[24]

The following tests allow for this problem by using series of errors measured as actual differences between past predictions and realizations, the latter being based exclusively on data that were available to participants in the successive surveys. The underlying argument is that the forecasters could and should have used this information so as to exploit and thereby eliminate as systematic elements in it. However, it must be noted that keeping track of the many successive revisions in complex data, particularly the quarterly national income and product accounts, is not a small or low-cost operation in which forecasters can be expected to engage routinely. The analysis that follows required creation of a comprehensive computer file of successive vintages of the data covered.[25]

Drawing upon that record to obtain the ex ante forecast errors as defined above, we next use these errors in autocorrelation functions of the general form

$$e_{t+j} = \sum_k^m \hat{\rho}_k e_{t+j-k} \quad , \quad k = j+1, \ldots, m \tag{13}$$

Here e_{t+j} represents the error of forecast made at time t for the j^{th} target quarter and $\hat{\rho}_k$ is the sample autocorrelation coefficients for the lag

---

[24]This is not to deny its validity as a criterion of statistical ex post assessments of the properties of the forecasts. The tests reported earlier in this paper can all be viewed as being of this nature.

[25]I am very much indebted to Louis Lambros for the accomplishment of this task.

k. The omission of $\hat{\rho}_k$ for $k \leq j$ reflects the fact that the information available at time t includes the errors of past predictions through the previous quarter $(t - 1)$ but does not include the errors of the current predictions for $t + j$.[26]

The autocorrelation functions (13) are computed for the errors in forecasts of those individuals who participated in more than 12 consecutive surveys, the same sample as that used before in the context of Table 4. Given these data, it seemed best to set $k = 6$ and again to use $j = 0, \ldots, 3$ (excluding Q4). The Box-Pierce statistic is then calculated by

$$Q_j = n(n + 2) \sum_k^6 (n - k)^{-1} \hat{\rho}_k^2 , \tag{14}$$

which is approximately distributed as chi-square, with $6 - j$ degrees of freedom.

Table 7 shows that the averages of the calculated Q_j values for the forecasts of GNP, RGNP, and CEDG decline systematically and strongly with the increase in the predictive horizon (column 2). The corresponding standard deviations show the same tendency but remain large in relative terms (column 3). For IPD, UR, and CBI, the mean values of Q are generally high and there is no evidence of any regular dependence of the distributions of the Q values on the distance to the target quarter.

The critical 10 percent level is widely used in practice as a cutoff for the Q test, and on this criterion most of the error series in most of the covered categories would pass the joint hypothesis that all of the examined

---

[26]For example, the errors of the Q0 forecasts will not be known until a quarter later, hence they are not yet available to the forecasts for Q1, Q2, and Q3, which are all made at the same time as those for Q0. The lack of current knowledge, then, impedes the elimination of significant autocorrelations for $\hat{\rho}_k$ where $k \leq j$. This argument applies here specifically to $\hat{\rho}_1$ for Q1, Q2, and Q3; $\hat{\rho}_2$ for Q2 and Q3; and $\hat{\rho}_3$ for Q3.

TABLE 7

CHI-SQUARE TESTS OF AUTOCORRELATIONS OF ERRORS IN SELECTED FORECASTS, SIX VARIABLES, 1968-1979

Quarter Predicted	Number of Forecast Series[a]	Q Statistics[b] Mean Value	Q Statistics[b] Standard Deviation	Percent of Forecasts with Q Statistics[b] That are Significant at the Level of 1%	5%	10%	Percent of Forecasts with $\hat{\rho}_k > 2/\sqrt{n}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Gross National Product (GNP)				
Q0	19	12.43	9.71	21.0	31.6	47.4	52.6
Q1	19	9.26	5.33	15.8	36.8	57.9	36.8
Q2	19	4.87	3.11	0	5.3	26.3	0
Q3	18	4.22	4.11	11.1	11.1	16.7	11.1
			Implicit Price Deflator (IPD)				
Q0	20	15.48	8.84	45.0	60.0	60.0	60.0
Q1	20	11.73	5.37	20.0	45.0	80.0	60.0
Q2	20	10.81	8.26	30.0	50.0	50.0	50.0
Q3	20	11.71	8.23	45.0	60.0	75.0	85.0
			GNP in Constant Dollars (RGNP)				
Q0	19	11.02	8.40	10.5	31.6	31.6	15.8
Q1	19	6.50	3.95	0	15.8	15.8	10.5
Q2	19	4.63	2.50	0	5.3	10.5	5.3
Q3	18	5.53	4.11	0	33.3	50.0	5.6
			Unemployment Rate (UR)				
Q0	18	9.20	6.83	22.2	27.8	27.8	22.2
Q1	18	9.46	5.87	22.2	33.3	38.9	33.3
Q2	18	9.00	8.80	33.3	44.4	44.4	55.6
Q3	17	9.26	10.64	23.5	52.9	58.8	47.1
			Consumer Expenditures-Durable Goods (CEDG)				
Q0	18	11.77	6.44	22.2	38.9	38.9	27.8
Q1	18	7.16	3.65	0	16.7	33.3	11.1
Q2	18	4.44	2.95	0	11.1	16.7	0
Q3	17	2.74	1.80	0	0	5.9	0

Table 7 (concluded)

Quarter Predicted	Number of Forecast Series[a]	Q Statistics[b] Mean Value	Standard Deviation	Percent of Forecasts with Q Statistics[b] That are Significant at the Level of 1%	5%	10%	Percent of Forecasts with $\hat{\rho}_k > 2/\sqrt{n}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Change in Business Inventories (CBI)							
Q0	20	10.76	7.10	15.0	35.0	45.0	60.0
Q1	20	9.57	5.73	10.0	45.0	45.0	35.0
Q2	20	11.44	8.42	25.0	60.0	65.0	80.0
Q3	20	6.31	4.94	15.0	40.0	45.0	20.0
Summary[c]							
Variable							
GNP	75	7.69	6.90	12.0	21.3	37.3	25.3
IPD	80	12.60	7.96	35.0	53.8	66.2	63.8
RGNP	75	7.00	5.74	2.7	21.3	26.7	9.3
UR	71	9.23	8.03	25.4	39.4	42.3	39.4
CEDG	71	6.49	5.28	5.6	16.9	23.9	9.9
CBI	80	9.52	6.85	16.2	45.0	50.0	48.8

[a]Includes forecasts of those individuals who participated in more than 12 consecutive surveys. The few observations available for Q4 are excluded. Base of the entries to the right.

[b]Refers to the Box-Pierce statistics as defined in eq.(14), with 6 - j degrees of freedom. See text.

[c]Covers all forecasts used in this table, summarized across the four target quarters, Q0-Q3.
SOURCE: Quarterly ASA-NBER surveys, 1968:4-1979:1.

autocorrelation coefficients are zero. The tests for RGNP and CEDG are the most favorable in this regard (see columns 4-6 and the summary in Table 7). However, two-thirds of the series for IPD and half of those for CBI have statistics that are significant at the 10% level. The frequencies of autocorrelated errors are also large for the short forecasts of GNP and the long forecasts of UR. Thus many forecasters appear to have failed to treat their own past errors efficiently as data to learn from, for one reason or another (inconsistent or deficient information, models, and judgments, surprisingly large and frequent disturbances).

It should be noted that these chi-square tests are neither strong nor direct.[27] An additional test is performed by inspecting all individual $\hat{\rho}_k$ coefficients to see how many of them fall outside of the range of two standard deviations from zero. The results, listed in the last column of Table 7, agree generally well with our earlier conclusions.

Evidence from the Group Mean Forecasts

Table 8 presents sample estimates of the autocorrelation functions (eq. 13) for the errors in the ASA-NBER group mean forecasts. If the error series, each of which contains 42 observations, were white noise, the standard deviation of $\hat{\rho}_k$ would be approximately 0.154. Of the 108 entries in columns 1-6 of the table, 82 are smaller than 0.154 in absolute value; 22 fall between 0.154 and 0.301; and only four exceed 0.301, that is, are outside the range of ± 2 s.d. from the mean zero. Inflation forecasts account for eight of the observations in the second and all four observations in the third group.

---

[27]For example, a value of Q below the 10% level indicates a probability of less than 90 percent that the hypothesis that the errors are not white noise is true. For more detail and examples, see Pindyck and Rubinfeld, 1981, pp. 549-550.

TABLE 8

TESTS OF AUTOCORRELATION OF ERRORS IN 24 SERIES
OF GROUP MEAN FORECASTS, 1968-1979

Quarter Predicted	Estimated Autocorrelation Coefficients[a]						Box-Pierce Statistic[b]
	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	$\hat{\rho}_5$	$\hat{\rho}_6$	Q_j
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
GNP in Current Dollars (GNP)							
Q0	-.18	-.15	-.04	-.06	-.17	.14	5.11
Q1		-.16	-.11	-.08	-.07	.11	2.84
Q2			-.05	-.02	.05	.08	.92
Q3				.09	.03	.10	.92
Implicit Price Deflator							
Q0	.35	.20	.23	.01	-.13	-.34	16.54#
Q1		.21	.22	.12	-.14	-.26	9.66§
Q2			.24	.11	-.17	-.32	10.14#
Q3				.12	-.20	-.41	11.09#
GNP in Constant Dollars (RGNP)							
Q0	.01	-.04	.01	-.10	-.18	.07	2.57
Q1		-.09	-.02	-.11	-.17	-.00	2.37
Q2			.01	-.02	-.09	-.05	.53
Q3				.03	-.06	-.09	.72
Unemployment Rate (UR)							
Q0	.04	-.18	-.05	-.12	-.15	.16	4.77
Q1		-.11	-.24	-.17	-.11	.04	5.53
Q2			-.22	-.20	-.09	.03	4.78
Q3				-.14	-.08	-.00	1.25
Consumer Expenditures--Durable Goods (CEDG)							
Q0	-.29	-.09	-.22	.19	.07	.12	9.07
Q1		-.15	-.14	.13	-.03	.12	3.52
Q2			-.12	.10	.04	.02	1.31
Q3				.13	-.00	.06	1.00
Change in Business Inventories (CBI)							
Q0	.11	-.02	-.09	.07	-.09	.05	1.70
Q1		-.02	-.12	.01	-.11	-.03	1.37
Q2			-.07	.02	-.11	-.04	1.01
Q3				.02	-.01	-.04	.12

[a]For level errors in UR and CBI, percentage change errors in the other variables. All measures refer to the means of predictions by those individuals who participated in al least 12 surveys. See equation 13 and text.

[b]See equation 14 and text. # means significant at the 5% level, § at the 10% level.

SOURCE: Quarterly ASA-NBER surveys, 1968:4-1979:1.

Not surprisingly, the Q statistics are definitely significant for the IPD errors, but the same does not apply to the other series, where they are actually rather small, with only a few exceptions (column 7). In several cases, the calculated Q's decline between Q0 and Q3, notably so for GNP and CEDG.

There is no indication that the absolute values $|\hat{\rho}_k|$ are systematically related to the lag k. In particular, they do not tend to decline as k rises (for IPD the $\hat{\rho}_6$ values, all negative, are particularly large). It is not clear that autocorrelations of higher order among the errors of these composite forecast series deserve much attention, but it certainly cannot be assumed that all or even most of them are zero.[28]

V. Summary and Conclusions

Main Results

1. The hypothesis that the regressions of actual on predicted values have zero intercepts and unitary slopes is rejected at the 5% significance level for 362 of the 2,350 forecast series examined (15.4%). Nearly half of these rejections refer to the inflation (IPD) forecasts, where they account for 44.3% of the regressions. The combined result for the otehr five variables is 187 rejections, or 9.6% of the 1,955 trials. I conclude that these weak tests of rationality are quite unfavorable to expectations of inflation, while showing other forecasts generally in much better light.

---

[28]In an earlier study based on ex post errors in the group mean forecasts and using as many as twelve autocorrelation lags, some of the $\hat{\rho}_k$ coefficients for k of 8, 9, and 10 quarters were found to be large and significant (see Zarnowitz, 1982b, Table 9 and text). However, one would expect the autocorrelations to be on the whole lower for the errors that are knowable ex ante than for the ex post errors, and the evidence we have tends to be consistent with that expectation.

attitudes of the University of Michigan Institute for Social Research (ISR). These questions have dealt mainly with the direction, not the size, of the expected price changes and they were altered repeatedly over the period, so that here the creation of a group forecast series requires a rather elaborate ex post procedure of quantitifying qualitative responses. Some of the studies find that the unbiasedness hypothesis cannot be rejected for the ISR data, others merely that it is "not so decisively rejected" as the inflation forecasts by economists and business executives.[29]

The regressions of actual on predicted inflation have also been found to produce serially correlated residuals, which some of the studies interpret as another departure from rationality. But the correctness of this view depends on the (generally unexamined) extent to which the calculated regression error terms constitute information knowable at the time of the forecast.

Tests for the joint null hypothesis of unbiasedness based on both OLS and GLS regression estimates are applied in McNees, 1978, to IPD, RGNP, and UR forecasts from three well-known econometric service bureaus, Chase, DRI, and Wharton. The periods covered are short, 5 $^1/_2$ or 6 years beginning in 1970:2, so the power of these tests is low, and the results are in part difficult to rationalize. For the multiperiod forecasts of inflation, the F statistics are generally significant but much higher for the GLS than the OLS estimates. For real growth, the situation is reversed and the null hypothesis is consistently

---

[29]For a comprehensive discussion of rationality tests with applications to the ISR data, see Huizinga, 1980; also Juster, 1979; Curtin, 1982; and Gramlich, as quoted. Business forecasts of price changes for goods and services sold and capital goods purchased come from the plant and equipment surveys of the Bureau of Economic Analysis, U. S. Department of Commerce; they have been examined by deLeeuw and McKelvey, 1981, and fail to pass the F test for unbiasedness decisively in 1970-80 as noted by Gramlich. Papadia, 1982, has applied the tests for aggregate results from consumer surveys conducted three times a year since 1973 or 1974 in seven EEC countries; he finds that the hypothesis of unbiasedness can be rejected in about half of the cases.

accepted for predictions over more than one quarter when GLS is used. The results for UR are quite mixed, with indications of bias in the predictions of cumulative change over the four-quarter span but not in the one-quarter ahead forecasts.

The first half of the 1970's was clearly among the most trying times for the forecasters generally (see Zarnowitz, 1979). But this is not to say that the forecast period somehow explains or excuses the observed failures of the forecasts to avoid bias and inefficiency. After all, it is precisely in times of highly variable inflation and real growth rates that the incentives to use data and predict efficiently are especially high. Moreover, as suggested by the present study, much of the variation among the forecasts is attributable to differences between the sources, models, variables, and horizons involved; it simply cannot be explained by differences in the periods covered.

References

Becker, Gary S. 1962. "Irrational Behavior and Economic Theory," Journal of Political Economy 70, no. 1 (February): 1-13.

Box, G. E. P. and Pierce, David A. 1970. "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models, Journal of the American Statistical Association 65, no. 332 (December): 1509-1526.

Carlson, John A. 1977. "A Study of Price Forecasts," Annals of Economic and Social Measurement 6, no. 1 (Winter): 27-56.

Cole, Rosanne. 1969. Errors in Provisional Estimates of Gross National Product. New York: National Bureau of Economic Research.

Curtin, Richard T. 1982. "Determinants of Price Expectations: Evidence from a Panel Study," in H. Laumer and M. Ziegler, eds., International Research on Business Cycle Surveys, Gower Publishing Company, Ltd.

deLeeuw, Frank and McKelvey, Michael. 1981. "Price Expectations of Business Firms," Brookings Papers on Economic Activity, Vol. 1: 299-313.

Figlewski, Stephen and Wachtel, Paul. 1981. "The Formation of Inflationary Expectations," Review of Economics and Statistics 63, (February): 1-10.

Gramlich, Edward M. 1983. "Models of Inflation Expectations Formation: A Comparison of Household, Economist, and Business Forecasts," forthcoming in Journal of Money, Credit, and Banking.

Howrey, E. Philip. 1978. "The Use of Preliminary Data in Econometric Forecasting," Review of Economics and Statistics 60, no. 2 (May): 193-200.

Huizinga, John. 1980. "Real Wages, Employment, and Expectations," Ph.D. thesis, M.I.T. Chapter 2.

Juster, F. Thomas. 1979. Statement presented in Hearings before the Task Force on Inflation of the Committee on the Budget, House of Representatives, 96th Congress, First Session (June), Washington, D.C.: U. S. Government Printing Office, pp. 214-226.

McNees, Stephen K. 1978. "The Rationality of Economic Forecasts," American Economic Review 68, no. 2 (May): 301-305.

Mincer, Jacob. 1969. "Models of Adaptive Forecasting," in J. Mincer, ed., Economic Forecasts and Expectations. New York: National Bureau of Economic Research: 81-111.

Mincer, Jacob, and Zarnowitz, Victor. 1969. "The Evaluation of Economic Forecasts," in J. Mincer, ed., Economic Forecasts and Expectations. New York: National Bureau of Economic Research: 3-46.

Moore, Geoffrey H. 1977. "Lessons of the 1973-1976 Recession and Recovery," in William Fellner, ed. Contemporary Economic Problems 1977. Washington, D. C.: American Enterprise Institute: 156-158.

Muth, John F. 1961. "Rational Expectations and the Theory of Price Movements," Econometrica 29, no. 3 (July): 315-335.

Nelson, Charles R. 1975. "Rational Expectations and the Predictive Efficiency of Economic Models," Journal of Business 48, no. 3 (July): 331-343.

Papadia, Francesco. 1982. "Rationality of Inflationary Expectations in the European Economic Communities' Countries," EEC Working Paper, Brussels (March) mimeo.

Pearce, Douglas K. 1979. "Comparing Survey and Rational Measures of Expected Inflation: Foercast Performance and Interest Rate Effects," Journal of Money, Credit, and Banking 11, no. 4 (November): 447-456.

Pindyck, Robert S., and Rubinfeld, Daniel L. 1981. Econometric Models and Economic Forecasts. New York: McGraw-Hill Book Co. (second edition).

Poole, William. 1976. "Rational Expectations in the Macro Model," Brookings Papers on Economic Activity 2: 463-514.

Theil, Henri. 1965. Economic Forecasts and Policy. Amsterdam: North-Holland (first edition 1958).

Wachtel, Paul. 1977. "Survey Measures of Expected Inflation and Their Potential Usefulness," in Joel Popkin, ed., Analysis of Inflation: 1965-1974. Cambridge, Mass.: Ballinger Publishing Company for the National Bureau of Economic Research, pp. 361-395.

Zarnowitz, Victor. 1969a. "The New ASA-NBER Survey of Forecasts by Economic Statisticians," American Statistician (February): 12-16.

Zarnowitz, Victor. 1974. "How Accurate Have the Forecasts Been?" in William F. Butler, Robert A. Kavesh, and Robert B. Platt, eds., Methods and Techniques of Business Forecasting. Englewood Cliffs, N.J.: Prentice-Hall: 565-596.

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. 1982a. "On Functions, Quality, and Timeliness of Economic Information," Journal of Business 55, no. 1 (January): 37-119.

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

Zarnowitz, Victor. 1982c. "The Accuracy of Individual and Group Forecasts from Business Outlook Surveys," NBER Working Paper No. 1053 (December).

Zellner, Arnold. 1979. "Posterior Odds Ratios for Regression Hypotheses: General Considerations and Some Specific Results," paper presented at the Econometric Society Meeting, Atlanta (November).