

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

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: Rational Expectations and Macroeconomic Forecasts

Chapter Author: Victor Zarnowitz

Chapter URL: <http://www.nber.org/chapters/c10387>

Chapter pages in book: (p. 462 - 491)

16 Rational Expectations and Macroeconomic Forecasts

16.1 Questions and Data

16.1.1 On Economics of Expectations and Surveys of Forecasts

Recent theorizing about expectations has concentrated on market rewards that motivate people to use all of the information that can be acquired cost-effectively. The rational expectations hypothesis (REH) assumes that a sufficiently large number of agents know “how the world works”; that is, they properly identify the structure of their environment and succeed in exploiting the existing profit opportunities. Their expectations are decisive for what transpires in the marketplace and are reflected in the equilibrating behavior of prices and other endogenous variables (Muth 1961; Poole 1976). Prices in a market may incorporate all of the information that matters, even though price expectations of many, perhaps even most, traders do not meet the rationality criterion. The competitive game of economic prediction cannot be comprehended by treating expectations as if they were single valued and universally shared; it is important to distinguish between *individual* and *market* expecta-

Reprinted from the *Journal of Business and Economic Statistics* 3 (October 1985): 293–311. © 1985 American Statistical Association.

This study is part of the National Bureau of Economic Research Program on Economic Fluctuations. Parts of an earlier version were presented at the annual meeting of the American Association for the Advancement of Science in Washington, D.C., on January 8, 1982, and before the Econometric and Statistical Colloquium at the University of Chicago. The author thanks the participants for useful comments. He is grateful to Craig Ansley, Edward George, John Huizinga, Robert Kohn, and Arnold Zellner for helpful discussions, and to the anonymous referees for criticisms that resulted in much improvement of the first submission. He is particularly indebted to Louis Lambros and Hanan Jacoby for valuable suggestions and research assistance and to Paul Flignor and Scott Katzer for help with computations. 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. The usual disclaimer applies.

tions. (For an early argument that rational market reactions may coexist with much individual "irrationality," see Becker 1962; a more recent discussion is in Mishkin 1981a.)

A model consisting of isolated competitive auction markets in which only price signals count is easy to conceive and manipulate, but of limited interest. In a modern economy, not all markets are of this type; uncertainty and a high degree of interdependence exist. Hence important signals are being transmitted by quantities as well as prices, and by global variables as well as local variables. Aggregate measures such as real GNP growth, inflation, unemployment, sensitive cyclical indicators, changes in money and credit, interest rates, and exchange rates are 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, for both public and internal uses, and many of them participate in periodic business outlook surveys. It might be argued that these are the *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 right. Indeed many of them are influential agents who have passed critical market tests, as certified by their positions and the rewards their forecasts and advice have earned 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). (For examples and further references, see Theil 1965; Mincer 1969b; Mincer and Zarnowitz 1969; Zarnowitz 1974, 1979; McNees 1978; Nelson 1975; Carlson 1977; Wachtel 1977; Pearce 1979; Figlewski and Wachtel 1981; and Gramlich 1983.) This article will adopt the first of these perspectives.

An ideal survey would use a large, properly constructed random sample to

ensure that the respondents represent well the universe of those whose expectations count and a system of rewards and penalties to ensure that the respondents 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 toward safeguarding against such pitfalls. This article benefits from the author's direct involvement with the management of the surveys to be discussed.

16.1.2 Tests of Rationality

The strictest interpretation of rational expectations is that a forecast $P_{ijt} = P_{i,t+j}$ equals the mathematical expectation of the corresponding target variable $A_{jt} = A_{t+j}$ conditioned on the relevant subset of the available information I (data and models). Here i refers to the source, t to the date, and j to the span or horizon of the forecast. Formally,

$$(1) \quad \forall i, t, j \exists I_{t-1}: P_{ijt} = E(A_{jt} | I_{t-1}), \quad i = 1, \dots, N; \\ t = 1, \dots, T; j = 0, \dots, S,$$

where E is the expected-value operator. There is no restriction on the span of the forecast (S may equal T), but the dating assumes for simplicity a one-period information lag. Thus the shortest forecasts ($j = 0$) are made in and for the period t , the next longer ones ($j = 1$) are made in t for $t + 1$, and so on, and all of these simultaneous multiperiod forecasts are based on I_{t-1} .

In (1) the information set I_{t-1} is taken to be common to all individuals. But if all were able and willing to use exactly the same data and the same models in the same way, then their simultaneous forecasts for any given target would be identical, that is, $\forall i, t, j: P_{ijt} = P_{jt}$. Presumably, I_{t-1} would then indeed contain all of the pertinent information generated in the economy. The forecasts would be unbiased, indeed tending to be self-fulfilling whenever A_{jt} depends on actions of economic agents based on P_{jt} . The forecast errors $\epsilon_{jt} = P_{jt} - A_{jt}$ would reflect solely the irreducible variation of the actual values; that is, P_{jt} would be optimal in the sense of having minimum error variances (the analog of perfect foresight in a stochastic world).

Both general reasoning and specific evidence indicate that this case is far from reality. Valuable information is limited and unevenly diffused, and its costs and returns to individuals with different skills and interests vary. There is uncertainty and disagreement as to the data and models to be used for the same predictive purpose. Some data are preliminary and subject to unpredictable revisions; some models contradict others, so not all can be true. Those forecasters who use better data and models will be less frequently misled, and

their expectations will be on the average more accurate. Furthermore, depending on the forecasting objective and the appropriate model, the criterion of unbiasedness may apply to some variables but not to others. Thus rationality does not mean the use of unbiased predictors in all cases (as shown in Grossman 1975).

Finally, even if forecasters use the same information for the same purposes, they may have different loss functions, in which case their forecasts would differ. For those whose loss functions are asymmetrical, the use of biased predictors may be optimal.

For all of these reasons, one would not expect individual forecasts to be generally uniform, unbiased, or self-fulfilling. Studies of survey data confirm the existence of substantial differences among concurrent and corresponding forecasts from different sources (Figlewski and Wachtel 1981; Jonung 1981; Zarnowitz 1984a).

If the reported forecasts are based on individual-specific, not general, information, then (1) is replaced by

$$(2) \quad \forall t, j \quad P_{ijt} = E(A_{jt} | I_{ijt-1}).$$

Here, for any given variable, the data and models used are those selected by the source, and the information may also vary for different predictive horizons.

Under (2), unlike under (1), the forecasts or expectations P_{ijt} cannot, of course, be all optimal; moreover, there is no longer any strong presumption that they will tend to be unbiased, with identically and independently distributed errors. People interact in making predictions and attempt to learn from their own errors and those of others. This process may be thought of as one of striving to reduce the gap between the individual and the general information sets, I_i and I , at any point in time. This means that each agent or forecaster tries to predict the average or prevailing forecast or what others are likely to predict that forecast will be. But this is a difficult, in principle open-ended, task. Individual rationality does not necessarily imply convergence of individual expectations and consistency of individual plans, at least not in the short-run reference frame of the forecasts under consideration (cf. Frydman and Phelps 1983).

It remains true that an individual forecaster's errors $\varepsilon_{ijt} = P_{ijt} - A_{jt}$ can satisfy any or all of the following criteria, each of which is desirable:

$$(3) \quad \forall t, j \quad E(\varepsilon_{ijt}) = 0,$$

$$(4) \quad \forall t, j \quad E(P_{ijt} \varepsilon_{ijt}) = 0$$

and,

$$(5) \quad \forall t, j, k \neq 0 \quad E(\varepsilon_{ijt} \varepsilon_{ijt-k}) = 0.$$

Forecast sets that meet all of the preceding criteria are accurate on the average or in the long run; they are uncorrelated with their own errors; and their errors

are independent in the sense of being serially uncorrelated. They may differ across sources in that some display larger, and others smaller, random errors, but all are free of systematic errors defined as failures to conform to (3)–(5).

One way to check on the extent to which genuine forecasts from surveys satisfy these criteria of rationality is by estimating the regression equations

$$(6) \quad A_{jt} = \alpha_{ij} + \beta_{ij}P_{ijt} + u_{ijt}$$

and then testing the joint null hypothesis $H_0: (\alpha_{ij}, \beta_{ij}) = (0, 1)$ for each target variable covered. If H_0 is accepted, $A_{jt} = P_{ijt} + u_{ijt}$ and $E(u_{ijt}) = 0$; that is, the forecasts are unbiased.

Some tests of H_0 assume that $E(u_{ijt}) = 0$, $E(u_{ijt}^2) = \sigma_{ijt}^2$, and $E(u_{ijt}u_{ijt-k}) = 0$ for $k \neq 0$. But it is possible for a forecast set to satisfy (3) or (4) or both but not (5), or vice versa. We must therefore examine separately the serial correlation and other properties of the forecast errors.

The advantage of this approach is that no specification is needed of what data and models the forecasters could and should have used; the only information required for the tests is matching time-series data on forecasts and realizations. As a rule, the outside observer has no way of knowing what the information set of a particular forecaster contains; indeed, even the producer of the forecast would probably often find it difficult to define the contents of that set clearly and exhaustively. (This statement is presumably more valid for judgmental than for econometric or time-series forecasts; but true ex ante predictions of all kinds virtually always include important judgmental elements.) On the other hand, the criteria used here are still weak, since rational expectations imply the efficient use of pertinent information, not just unbiasedness and serially uncorrelated errors. The forecaster's own past errors are only a part of his or her information set.

16.1.3 Sources of Evidence, Definitions of Measures, and Scope of Study

Owing to the efforts of the National Bureau of Economic Research (NBER), in collaboration with the American Statistical Association (ASA), a large amount of information has been assembled on the record of forecasting changes in the U.S. economy. The data are believed to represent well the contemporary state of the art in this area.

Each quarter, the NBER examines the results of a questionnaire mailed by the ASA. (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 since 1974 in *AmStat News*. The forecasts have also been published and discussed in *Economic Outlook USA*, a report by the Survey Research Center at the University of Michigan. On the origin of the survey and the design of the questionnaire, see Zarnowitz 1969c.) The survey reaches a broadly based and diversified group 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 11 time series representing the principal measures of national output, income, consumption, investment, the price level, and unemployment. (Since 1981:2 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 1984a.)

Past work on the survey data has concentrated on summary measures, group medians, or means, whereas this article is part of a comprehensive study of forecasts by individual respondents in the ASA-NBER group. Furthermore, unlike the many recent studies that consider only expectations of inflation, this report covers other important aggregate 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). Many individuals responded only once or a few times, however, and some decision had to be made on the minimum number of survey responses that would qualify a participant for inclusion. It was set at 12, which left 79 individuals in the sample. (There are a few exceptions where a series contains less than 12 observations. These refer to the longer horizons and arise because some forecasters occasionally predicted fewer than four quarters ahead. For example, of the 395 \dot{P}_{ij} series for GNP, 16 [4%] have 10 or 11 observations each, all but four of them for Q4.)

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 that reflect the real growth or inflation or both. These are gross national product (GNP) and consumer expenditures for durable goods (CEDG), both in current dollars; 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 ($\dot{P}_{ijt} - \dot{A}_{ijt}$). The definitions employed are

$$(7) \quad \dot{A}_{ijt} = \left(\frac{A_{i,t+j} - A_{i,t+j-1}}{A_{i,t+j-1}} \right) 100, \quad j = 0, \dots, 4,$$

and

$$(8) \quad \begin{aligned} \dot{P}_{ijt} &= \left(\frac{P_{it} - A_{it-1}^*}{A_{it-1}^*} \right) 100, & \text{if } j = 0 \\ &= \left(\frac{P_{i,t+j} - P_{i,t+j-1}}{P_{i,t+j-1}} \right) 100, & \text{if } j = 1, \dots, 4. \end{aligned}$$

The change in business inventories (CBI), a current-dollar series, is in first-difference form and does not appear to have a trend. The unemployment rate (UR) represents the percentage of the civilian labor force that is unemployed 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, that is, predicted level minus actual level ($P_{jt} - A_{jt}$). (For further references to the treatment of level and change errors, see Zarnowitz 1984a.)

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 preceding quarter, which are marked $A_{j,t-1}^*$ in (8). (An exception is the UR series, which is available monthly.) 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, that 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. It should be noted that this procedure imposes on the forecasters the burden of predicting 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 article. (On the role of preliminary data and revisions in economic measurement and prediction, see Cole 1969b; Howrey 1978; and Zarnowitz 1979, 1982c.)

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.

In all, 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 79 individuals and the group averages). The volume and quality of the data are such as to permit an intensive study of several aspects of the economic predictions. 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: Carlson 1977; Pearce 1979; and Figlewski and Wachtel 1981.) The problems relate to the timing of the survey, its consistency, and the effective forecasting spans involved; changes in composition over time; the role of outliers; and reporting errors. A careful proofreading of the data is needed to detect simple mistakes

of calculation, copying, and typing that chance or neglect will always occasion in some replies. The voluminous ASA-NBER materials were submitted to such an audit with the aid of a computer and, where needed, inspection of the original submissions. Although the number of mistakes identified by this means turned out to be very small in relative terms, failure to eliminate them would have adversely affected the evaluation of several individual records.

Section 16.2 presents the evidence on the question of bias in multiperiod predictions by individuals and groups. Section 16.3 addresses the problem of serially dependent residual errors and discusses tests for autocorrelation in the "knowable" forecast errors. Finally, section 16.4 sums up the results and places them in the context of earlier related work.

16.2 Testing for Bias in Multiperiod Predictions

16.2.1 Distributions of the Regression and Test Statistics

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

$$(9) \quad \dot{A}_{jt} = \alpha_{jt} + \beta_{jt} \dot{P}_{ijt} + u_{ijt}, \quad j = 0, \dots, 4; t = 1, \dots, T.$$

For the unemployment rate (UR) and inventory investment (CBI), levels A_{jt} and P_{ijt} were used instead. Estimation of the regressions requires certain assumptions about the probability distribution of the disturbances u_{ijt} . The simplest and most common approach is to assume that $E(u_{ijt}) = 0$, $\text{var}(u_{ijt}) = \sigma_{ij}^2$, and the u_{ijt} 's are uncorrelated for each forecaster and target variable covered. The technique of ordinary least squares (OLS) applies in this case. The sample least squares estimates $\hat{\alpha}$ and $\hat{\beta}$ (the subscripts may now be dropped for simplicity) lend themselves to statistical tests of the joint null hypothesis 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 forecasts contain some systematic errors.

Table 16.1 represents the evidence from the full collection of 790 P_{ij} and 1,560 \dot{P}_{ij} forecast series. To provide a background of descriptive statistics, the OLS estimates of the intercepts and slope coefficients are summarized in columns 1-4.

The means of $\hat{\alpha}_{ij}$ ($\bar{\alpha}$) tend to increase with j , the distance to the target quarter, at least from Q0 through Q3, for most variables (col. 1). In contrast, the means of $\hat{\beta}_{ij}$ ($\bar{\beta}$) typically decrease (col. 3). The standard deviations of $\hat{\alpha}_{ij}$ and $\hat{\beta}_{ij}$ both tend to rise as the predictive horizon lengthens, with only a few exceptions (cols. 2 and 4). These measures indicate a great deal of dispersion, reflecting partly differences in the ability of the individuals to produce un-

Table 16.1 Sets of Multiperiod Forecasts for Six Aggregate Variables, 79 Participants in ASA-NBER Surveys, 1968-79: Distributions of Regression Statistics and Tests of Bias

Quarter Predicted	Mean Values of Individual Statistics ^a				Percentage of Forecasters with Significant Tests ^b				
	$\bar{\alpha}$ (1)	S.D. _{α} (2)	$\bar{\beta}$ (3)	S.D. _{β} (4)	F (5)	t_{α} (6)	t_{β} (7)	$F(s)$ (8)	$F(l)$ (9)
<i>GNP in current dollars (GNP)</i>									
Q0	.38	.76	.87	.32	12.7	15.2	12.7	0	21.7
Q1	.81	.75	.65	.34	10.1	17.7	19.0	3.0	15.2
Q2	1.17	.88	.52	.41	11.4	26.6	17.7	3.0	17.4
Q3	1.27	1.06	.46	.51	16.5	26.6	24.1	3.0	26.1
Q4	1.12	1.24	.50	.59	11.4	20.2	16.5	0	19.6
<i>Implicit price deflator (IPD)</i>									
Q0	.42	.42	.81	.33	26.6	19.0	17.7	11.8	37.8
Q1	.72	.51	.69	.42	46.8	36.7	17.7	20.6	66.7
Q2	1.03	.48	.48	.42	57.0	48.1	20.3	23.5	82.2
Q3	1.20	.46	.36	.41	64.6	43.0	17.7	20.6	97.8
Q4	1.27	.66	.42	.65	58.2	38.0	16.5	8.8	95.6
<i>GNP in constant dollars (RGNP)</i>									
Q0	-.12	.36	1.06	.31	10.1	19.0	12.7	2.9	15.6
Q1	-.29	.48	1.04	.43	8.9	7.6	8.9	0	15.6
Q2	-.11	.65	.80	.59	8.9	7.6	8.9	0	15.6
Q3	.01	.72	.62	.64	12.7	2.5	7.6	0	22.2
Q4	-.27	1.19	.72	1.04	15.2	0	12.7	0	26.7
<i>Unemployment rate (UR)</i>									
Q0	-.01	.26	1.00	.05	2.5	3.8	3.8	0	4.3
Q1	-.01	.76	1.01	.14	2.5	2.5	3.8	0	4.3
Q2	.29	1.17	.98	.22	3.8	3.8	3.8	0	6.5
Q3	1.01	1.38	.88	.26	12.7	7.6	8.9	3.0	19.6
Q4	1.80	2.12	.75	.39	10.1	16.5	11.4	0	17.4
<i>Consumer expenditures for durable goods (CEDG)</i>									
Q0	.99	.87	.93	.45	20.0	12.0	13.3	6.5	29.5
Q1	1.26	1.16	.43	.55	6.7	8.0	16.0	0	11.4
Q2	1.55	1.16	.27	.67	8.0	8.0	10.7	3.2	11.4
Q3	1.41	1.70	.26	.82	2.7	2.7	13.3	3.2	2.3
Q4	.57	1.88	.59	.92	4.0	5.3	4.0	0	6.8
<i>Change in business inventories (CBI)</i>									
Q0	2.76	3.54	.88	.52	16.2	20.3	8.9	11.8	19.6
Q1	1.81	4.54	.93	.61	10.0	13.9	13.9	2.9	15.2
Q2	2.18	5.61	.82	.80	6.3	10.1	5.1	5.9	6.5
Q3	2.22	5.72	.78	.74	3.8	6.3	3.8	0	6.5
Q4	2.97	4.84	.78	.59	3.8	6.3	2.5	2.9	4.3

^aThe entries in cols. 1 and 2 are the means ($\bar{\alpha}$) and standard deviations (S.D. _{α}) of the $\hat{\alpha}_j$ estimates from the regressions of actual values on the individual forecasts. The entries in cols. 3 and 4 are the means ($\bar{\beta}$) and standard deviations (S.D. _{β}) of the $\hat{\beta}_j$ estimates from the same regressions. See text and eq. (9). The regressions are estimated by OLS. All figures refer to those individuals who participated in at least 12 surveys: 75 for CEDG, 79 for each of the other variables.

Table 16.1 Continued

^bThe significance level is 5% for all tests. The percentages in cols. 5–7 refer to all participants in at least 12 surveys (same coverage as in cols. 1–4); col. 8, to those who responded to 12–19 surveys (31–34); and col. 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 $\beta = 1$. If the null hypothesis is true, the statistic for the joint test should have an F -distribution with 2 degrees of freedom in the numerator and $n - 2$ df in the denominator (where n , the number of observations, varies across the individuals). The appropriate t -tests are two-tailed, using the t -distribution with $n - 2$ df.

biased forecasts, partly differences in coverage (the samples are smaller for Q4 than for the other horizons), and partly differences between the variables (forecasts differ sharply across the individuals for volatile series that are especially difficult to predict, such as CBI).

A sharp contrast between the predictions of inflation and those of the other variables is shown by the relative frequencies of the individual forecast series that failed to pass the joint test for unbiasedness at the 5% significance level (col. 5). For IPD, about half of the computed F -ratios exceed the critical $F_{.05}$ values, whereas for GNP and RGNP the corresponding frequencies are 12% and 11%, respectively; for UR, CEDG, and CBI, they were 6%–8%.

According to the separate t -tests for regression intercepts and slopes, the incidence of $\alpha \neq 0$ is much higher for IPD than for GNP, whereas the incidence of $\beta \neq 1$ is similar for the two variables (cols. 6 and 7). These tests suggest that the high F -ratios observed for the inflation forecasts are associated to a larger extent with the deviations of α from 0 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 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 (cols. 5–7). The frequencies for UR generally tend to rise, whereas 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.

16.2.2 The Effects of Sample Size

Although 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 much shorter than would be desirable (inevitably so, since our surveys began only in 1968). The minimum requirement of participation in at least 12 surveys improves the data by eliminating the occasional respondents. 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 probabilities of wrongly rejecting the null hypothesis H_0 when it is true but also high probabilities of wrongly accepting H_0 when it is false. For small sample sizes, therefore, these tests have very low power against the alternative composite hypothesis that is merely a negation of H_0 (i.e., $H_1: \alpha \neq 0, \beta \neq 1$). As shown in Zellner 1984, several issues arise at this point: 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 almost habitually disregarded in applied economic and econometric literature.

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 12–19 surveys than for those who participated in 20 or more surveys (table 16.1, cols. 8 and 9). Indeed, the proportions for the first subset, $F(s)$, are typically 0 or less than 5% and average 1.9%, except for IPD, where they range between 9% and 24% and average 17.1%. In contrast, the proportions for the second subset, $F(l)$, are concentrated between 10% and 25% and average 14.4%, except again for IPD, where they range between 38% and 98% 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 the large discrepancies between the reported results for $F(s)$ and $F(l)$ cannot be attributed to differences in the periods covered. 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, the corresponding figures are 48 (8) and 38 (8), respectively. Thus some attrition occurred in the number of forecasters per survey. Its effect was about the same, however, 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.

To increase power, higher significance levels may be employed. Table 16.2 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 0, misleadingly suggesting that no bias at all exists in this group of relatively short forecast series (cf. cols. 2, 5, and 8). For the longer series, however, the decision to use 10% instead of 5% as the significance level would have made little difference in these conclusions, and even at the 1% level the negative results on the inflation forecasts are very evident in the $F(l)$ entries (cols. 3, 6, and 9).

Table 16.2 Summary of Results for Tests of $H_0: \alpha = 0, \beta = 1$; Two Groups of Forecasters, Six Variables, 1968–79

Variable	Percentage of Forecasts with F -Ratios That Are Significant at the								
	1% Level			5% Level			10% Level		
	F (1)	$F(s)$ (2)	$F(l)$ (3)	F (4)	$F(s)$ (5)	$F(l)$ (6)	F (7)	$F(s)$ (8)	$F(l)$ (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	.6	19.1	20.5	8.2	29.8
UR	0.5	0	0.9	6.3	.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	.8	0	1.3	8.0	4.7	10.4	17.2	14.7	19.1

Note: The abbreviations for the variables are identified in table 16.1. The entries in cols. 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 cols. 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 cols. 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).

16.2.3 Confidence Regions

Consider the ratio

$$(10) \quad F = [1/2(c_{11}c_{22} - c_{12}^2)s_u^2][c_{11}(\hat{\beta} - \beta)^2 + c_{22}(\hat{\alpha} - \alpha)^2 - 2c_{12}(\hat{\alpha} - \alpha)(\hat{\beta} - \beta)],$$

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 $\hat{\alpha}$ and $\hat{\beta}$ is given for any selected confidence coefficient g by $F \leq F_g$, where the probability $\Pr(F < F_g) = g$. It is an ellipse centered at $(\hat{\alpha}, \hat{\beta})$, and the higher g , the larger the ellipse ($g = .99, .95, .90$ correspond to the significance levels of 1%, 5%, and 10%, respectively, in the tests of H_0). In the present context, it is of interest to compare the confidence regions for selected short and long series of forecasts.

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, figure 16.1 shows the concentric ellipses associated with the confidence levels of .99, .95, and .90. The ellipses for the shorter series are much larger and spaced much more widely apart than those for the longer series (about twice as long and twice as wide, measured by the major and minor axes). As the figure demonstrates (and other such comparisons would generally confirm), the conventional use of high g

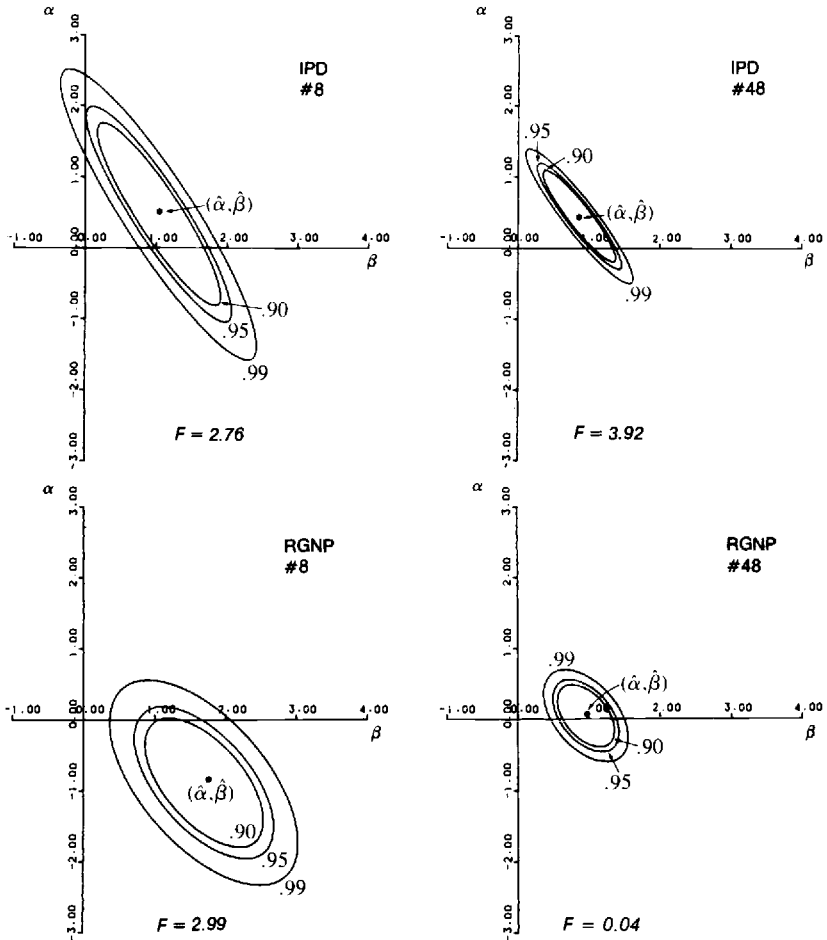


Fig. 16.1 Confidence regions for selected forecasts of inflation (IPD) and real growth (RGNP).

Note: 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 are listed in the figure.

values in analyzing small sets of predictions can be quite costly in terms of the precision sacrificed to reduce the probability of type I errors.

The high incidence of bias in the inflation forecasts is on the whole reaffirmed by this analysis (note the F values in fig. 16.1 and the location of the $[0, 1]$ points relative to the boundaries of the confidence regions). That the ellipses have downward-sloping major axes indicates that $\hat{\alpha}$ and $\hat{\beta}$ are negatively correlated, which simply reflects the fact that the mean values of the forecasts are positive. (An elementary property of the two-variable regression

model is that $\text{cov}(\hat{\alpha}, \hat{\beta}) = -\bar{x} \text{var } \hat{\beta}$, where \bar{x} is the mean of the explanatory variable. In the preceding regressions the forecasts play the role of x .)

16.2.4 Mean Errors

The tests summarized in tables 16.1 and 16.2 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, which has long been observed in a great variety of forecasts, can arise in unbiased as well as biased predictions (Mincer and Zarnowitz 1969, pp. 15–20). Table 16.3 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 quan-

Table 16.3 Selected Statistics on the Distribution of Mean Errors in Individual Forecasts, 1968–79

Quarter Predicted	Variables Predicted					
	GNP (1)	IPD (2)	RGNP (3)	UR (4)	CEDG (5)	CBI (6)
<i>Means (standard deviations) of the mean errors^a</i>						
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 (.24)	-.29 (.23)	.03 (.76)	-1.10 (3.10)
Q4	-.08 (.29)	-.61 (.21)	.53 (.31)	-.32 (.27)	.15 (.80)	-1.85 (2.80)
<i>Percentage of under (over) estimates^b</i>						
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)

^aThe errors are defined as the predicted minus the 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 $\bar{P}_{ij} - \bar{A}_{ij}$; 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 eqs. (7) and (8) for definitions of P_{ij} , A_{ij} , \bar{P}_{ij} , and \bar{A}_{ij} ; the overbars indicate averaging over time t .) The means of the mean errors across the individuals are without, the corresponding standard deviations are within, the parentheses.

^bThe 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).

tity component tend to cancel each other in the predictions of rates of change in the current-dollar GNP, where the mean errors are negative for most individuals but on the average very small throughout (cf. table 16.3, cols. 1–3). Underpredictions prevail for the unemployment rate in Q2–Q4 (consistent with the overprediction of real growth) and for business inventory investment, whereas the record for the rates of change in consumer durables is more mixed (cols. 4–6).

16.2.5 Estimates and Tests for Group Mean Forecasts

Consider now the regressions of the actual values on the series of means of the corresponding predictions by all individuals included in this study. Table 16.4 shows that the results vary greatly across our 30 different categories. The absolute values of the regression intercepts $|\hat{\alpha}|$ often increase with the predictive horizon, whereas the signs of these estimates are about equally mixed (col. 1). All of the slope coefficients ($\hat{\beta}$) are positive, but they otherwise display no common regularities (col. 2).

For GNP, the values of α do not deviate significantly from 0 nor the values of β from unity, according to the F - and t -ratios (cols. 3–5). In contrast, the F -tests strongly reject H_0 for the inflation forecasts, particularly in the more distant quarters, and the t -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 it is the short predictions 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 $\beta > 1$.

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 UR, which helps to explain why the r^2 coefficients are so much higher for the latter, 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 far more variable rates of change in CEDG beyond Q0. In general, bias does not imply particularly large errors, and some of the forecasts that appear to be highly biased are indeed relatively accurate (notably for UR, but also for the short predictions of IPD, RGNP, and CBI).

The Durbin-Watson (DW) statistics listed in table 16.4, column 7, suggest that the residual disturbances from the regressions of actual on predicted 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 strongly indicate 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 standard errors and upward the 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 these tests, which motivates some subsequent analysis (sec. 16.3.2).

16.2.6 Pooling the Forecasts

Let all individual series of forecasts for a given variable and span be stacked into one $NT \times 1$ vector \mathbf{P} . Let \mathbf{A} denote the $NT \times 1$ vector for the corresponding actual values, which includes N terms A_t in each of the T unit periods to match an equal number of predictions P_{it} , $i = 1, \dots, N$. Then the simplest model that combines time-series and cross-sectional data from business outlook surveys such as that of ASA-NBER can be written as

$$(11) \quad \mathbf{A} = (\mathbf{1} \mathbf{P})\boldsymbol{\gamma} + \mathbf{u},$$

where $\boldsymbol{\gamma}' = (\alpha \beta)$, $\mathbf{1}$ is an $NT \times 1$ vector of 1's, and \mathbf{u} is an $NT \times 1$ vector of residual errors. In (11) the parameters α and β are assumed to be fixed and common to all forecasters. With the data used in this article, each of such regressions would include some 1,600–1,800 observations. Such pooling might seem to offer a way to obtain more powerful and precise tests of the REH.

Figlewski and Wachtel (1981) used a regression like (11) to relate actual inflation to pooled individual forecasts of inflation from the semiannual Livingston survey (1,864 observations). They first estimated the equation by OLS and then reestimated by weighted least squares to allow for large differences in the residual variance over time. (In the weighted regression each observation is divided by the standard error of the residuals for the given time period as calculated by OLS.) They reported both regressions but not the test statistics, noting merely that “the F -test clearly indicates forecast bias” (p. 3) for the results.

The model under consideration is seriously misspecified, however. In each unit period, the dependent variable A_t takes on a single value, whereas the independent variable P_{it} takes on N values. Thus for any two forecasters k and l ,

$$(12) \quad u_{kt} - u_{lt} = \beta(P_{lt} - P_{kt})$$

in each of the T periods. Assuming that $\beta \neq 0$ and $P_{kt} \neq P_{lt}$ (to exclude entirely implausible and uninteresting cases), the differences among the errors mirror the differences among the predictions. For any given target in any single period, say A_1 for $t = 1$, P_{i1} and u_{i1} ($i = 1, \dots, N$) have equal variances and a correlation of -1 . This cross-sectional dependence of contemporaneous predictions and errors will cause $\text{cov}(P_{it}, u_{it})$ to be nonzero in the pooled model (11), even if for the individual forecast series the condition that the P 's and u 's are independent holds. In other words, the estimate of β in (11) is biased away from unity.

Table 16.4 Regressions of Actual on Predicted Values and Tests of Bias, 30 Series of Group Mean Forecasts, OLS Estimates, 1968-79

Quarter Predicted	Regression Estimates ^a		F-Ratio ^b for $\alpha = 0, \beta = 1$			t-Tests ^c for $\alpha = 0, \beta = 1$		Squared Correlation ^d		Durbin-Watson ^e	Standard Error (8)	Additional Statistics ^f	
	$\hat{\alpha}$ (1)	$\hat{\beta}$ (2)	$\alpha = 0, \beta = 1$ (3)	$\alpha = 0$ (4)	$\beta = 1$ (5)	(r^2) (6)	(7)	Mean (9)	S.D. (10)				
<i>Gross national product (GNP)</i>													
Q0	-.42	1.26	2.17	-1.28	1.72 [§]	.63	2.36 [°]	.64	2.23	1.06			
Q1	-.03	1.06	.28	-.05	.24	.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			
<i>Implicit price deflator (IPD)</i>													
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			
<i>Real gross national product (RGNP)</i>													
Q0	-.36	1.44	6.02*	-2.70*	3.45*	.75	2.47 [±]	.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			

<i>Unemployment rate (UR)</i>										
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
<i>Consumer expenditures for durable goods (CEDG)</i>										
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	2.52 ⁺	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
<i>Change in business inventories (CBI)</i>										
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

^aThe 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.

^bThe critical values of $F(2, 40)$ at the indicated significance levels are: 6.04 (.5%), 5.18 (1%), 4.06 (2.5%), 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.

^cFor the meaning of symbols *, °, and †, see n. b.

^dCorrected for df.

^eCorrected 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; †, result indeterminate; and +, positive serial correlation present. The 5% significance points of d_k and d_k' for $n = 40$ and $k = 1$ are 1.44 and 1.54, respectively.

^fStandard error of the regression: $\sqrt{\sum \mu^2/n - 2 \cdot S.D.}$; standard deviation of the dependent variable (actual values).

The misspecified error structure of (11) makes this a basically inappropriate model. The least squares estimates of α and β are biased even in large samples when $\text{cov}(P, u) \neq 0$. This is clear from an evaluation of their probability limits: $\text{plim } \hat{\beta} = \beta + [\text{cov}(P, u)/\text{var}(P)]$. If $\text{cov}(P, u) < 0$, as seems likely in view of (12), then the slope estimates from (11) will be biased downward, and the intercept estimates upward. Table 16.5 provides a few illustrations. A comparison of the corresponding entries in columns 1 and 2 of tables 16.5 and 16.4 shows that the estimates of α are much higher and those of β much lower in the pooled regressions than in the group mean regressions. (The same is true of the estimates reported by Figlewski and Wachtel: compare their eq. [3] for the average forecasts with their OLS pooled regression in footnote 6.) The standard errors are extremely low (the t -ratios high) in table 16.5, and the calculated F -values are significant even at the .1% level. For the short forecasts of UR, a is very close to 0 (.4) and b to 1 (.99), yet F is still as high as 34. The results in table 16.4 are far more favorable. The examples are typical and will suffice.

Suppose that the roles of actuals and forecasts are reversed as in

$$(13) \quad \mathbf{P} = (\mathbf{t} \mathbf{A})\delta + \mathbf{v},$$

where $\delta' = (\alpha' \beta')$. Here for any given target value such as A_1 , there is a distribution of the individual predictions P_{it} with $E(P_{it}) = \alpha' + \beta'A_1$ and $\text{var}(P_{it}|A_{1,t}) = \text{var}(v_{it})$. If $(\alpha', \beta') = (0, 1)$, the mean forecasts are unbiased. Indeed, the group mean forecasts from a series of surveys are on the average

Table 16.5 Pooling the Survey Forecasts: Some Examples

Quarter Predicted	$\mathbf{A} = (\mathbf{t} \mathbf{P}) \boldsymbol{\gamma} + \mathbf{u}$ (Eq. [11]) ^a					$\mathbf{P} = (\mathbf{t} \mathbf{A})\delta + \mathbf{v}$ (Eq. [13]) ^b	
	a (1)	b (2)	\bar{r}^2 (3)	F (4)	n (5)	a' (6)	b' (7)
<i>Gross national product (GNP)^c</i>							
Q0	.58 (.05)	.77 (.02)	.39	66.6	1,819	.99 (.04)	.50 (.01)
Q3	1.49 (.08)	.35 (.03)	.06	181.0	1,694	1.74 (.04)	.16 (.02)
<i>Implicit price deflator (IPD)^c</i>							
Q0	.62 (.03)	.67 (.02)	.37	228.1	1,808	.49 (.03)	.56 (.02)
Q3	1.34 (0.4)	.25 (.03)	.04	786.7	1,685	.85 (.04)	.17 (.02)
<i>Unemployment rate (UR)^d</i>							
Q0	.04 (.02)	.99 (.004)	.97	34.0	1,796	.12 (.02)	.99 (.004)
Q3	1.39 (.11)	.80 (.02)	.49	127.9	1,673	2.06 (.09)	.60 (.02)

^aCols. 1 and 2: OLS estimates of $\boldsymbol{\gamma} = (\alpha, \beta)$ and their standard errors (in parentheses). Col. 3: squared correlation coefficient corrected for df. Col. 4: F -ratio for $(\alpha, \beta) = (0, 1)$ with 2 and $n - 2$ df. Col. 5: number of observations.

^bOLS estimates of $\delta' = (\alpha', \beta')$ and their standard errors (in parentheses).

^cPercentage changes.

^dLevels.

over time more accurate than most of the sets of component individual predictions, though they are not always unbiased (see chapter 15 and preceding text). Model (13) yields answers to some interesting questions; for example, do forecasters perform better for some values of A than for others? This model, too, however, seems essentially unsuited for our present purpose. First, one of the proper criteria of our tests, namely, the independence of predictions and errors, is directly contradicted here, since P is a function of v . Second, it is natural to think of forecasts as being available before the unknown stochastic realizations, which suggests choosing P as the independent and A as the dependent variable, not the other way round. (Implicitly, if the possibility of *both* [11] and [13] is admitted, P , as well as A , is being considered a random variable. On the debate about the alternative interpretations and applicability of “classical” vs. “bivariate” regression, see Maddala 1977, pp. 97–101.)

The estimates of α' and β' are no more informative than those of α and β as far as the tests of our basic hypothesis are concerned. (See table 16.5, cols. 6 and 7 and compare the corresponding entries in cols. 1 and 2.) Note that \bar{r}^2 and N are the same for equations (11) and (13). Furthermore, $(\alpha, \beta) = (0, 1)$ for (11) implies that

$$(\alpha', \beta') = \left[\bar{P} \left(1 - \frac{\text{var}(P)}{\text{var}(A)} \right), \frac{\text{var}(P)}{\text{var}(A)} \right]$$

for (13), and the F -tests calculated either way on this basis lead to the same conclusions. The quite different null hypothesis that $(\alpha', \beta') = (0, 1)$ is also strongly rejected by the data.

In short, pooling macroeconomic survey forecasts of this type cannot be properly done within the general regression framework discussed earlier. Allowing the parameters of such models to vary across individuals and time, though in principle desirable, would not resolve this problem. Rather it should be instructive to work directly with forecast errors, using variance decomposition methods to isolate and study time as well as individual effects; but this is a task best left for another article.

16.3 Testing for Dependencies in Errors

16.3.1 Autocorrelated Disturbances and Bias

Tests for serial correlations among the regression residuals u_{ijt} (eq. [9]) have been made for all series that are sufficiently long and complete. These data refer to the forecasts by 18–20 individuals (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. There is adequate coverage

here for spans Q0–Q3; the samples for Q4, which are smaller, are not included.

The serial correlation coefficients are defined, on the assumption of homoscedasticity, as

$$(14) \quad \hat{\rho}_k = \widehat{\text{cov}}(u_t, u_{t-k}) / \widehat{\text{var}}(u_t), \quad k = 1, \dots, 6.$$

The Ljung-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

$$(15) \quad Q = n(n+2) \sum_k^6 (n-k)^{-1} \hat{\rho}_k^2,$$

which is approximately distributed as chi-square with six degrees of freedom (df). (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 $Q = \sum_k^m \hat{\rho}_k^2 \sim \chi_m^2$. In view of the small size of the available samples, it seemed advisable to avoid these common approximations. See Box and Pierce 1970.)

Most of the Q -statistics computed for the inflation and unemployment forecast series are found to be statistically significant at the 5% and 10% levels, and the frequencies are particularly high for IPD (see table 16.6, cols. 2–4). In contrast, the other variables suggest a relatively moderate incidence of autocorrelated residual errors. 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.

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 16.6, cols. 5–8). Although gaps in the data impair the comparability of the measures underlying this cross-tabulation, the broad indications obtained are of sufficient interest. Serial correlation in the error terms may produce large values of the F -statistics, wrongly causing them to reject the null hypothesis. For two thirds of the inflation (IPD) forecast series, both Q and F are significant, but elsewhere such cases are infrequent (col. 5), and those in which neither statistic is significant abound (col. 8). Except for the IPD and UR sets, the F 's are likely to be more significant when the Q 's are not (cf. cols. 5 and 7). Autocorrelation of error terms without bias is also often found, most prominently so for UR (col. 6).

16.3.2 Generalized Least Squares Estimates

Models with autoregressive errors of the form

$$(16) \quad A_t = \alpha + \beta P_{gt} + u_t \quad \text{and} \quad u_t = e_t - \sum_{i=1}^j \rho_i u_{t-i}$$

Table 16.6 Frequencies of Significant Q - and F -Statistics for Selected Forecast Sets, Six Variables, 1968–79

Variable	No. of Forecast Series ^a (1)	Significance Level of Q -Statistics ^b			Significance at the 10% Level ^c			
		1% (2)	5% (3)	10% (4)	Q and F (5)	Q only (6)	F only (7)	Neither (8)
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

Note: Cols. 2–8 are percentages based on entries in col. 1.

^aEach series includes consecutive observations only, by participants in more than 12 surveys.

^bRefers to the Box-Pierce statistics as defined in eq. (15) with 5 df. See text.

^cRefers 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 cols. 5–8 is 100.0.

were estimated for the mean forecast series P_{gr} . Here e_j is a normally and independently distributed error term and j equals 1, 2, and 3 for Q1, Q2, and Q3, respectively. (The procedure used is AUTOREG [SAS/ETS User's Guide, 1980 edition, pp. 8.1–8.7]. It is not applicable to data with missing values, hence the exclusion of Q4 from table 16.7.) Table 16.7 shows that 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. The generalized least squares (GLS) and OLS estimates also show no significant differences for the forecasts of RGNP in Q2 and Q3 and those of CEDG. If $\alpha = 0$ and $\beta = 1$, the forecast errors would optimally follow moving-average (MA) processes of the order $j - 1$ (Granger and Newbold 1977, pp. 121–22). The results reported in table 16.7, columns 6–8, for Q3 and perhaps Q2 are suggestive of the presence of such processes.

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. Still, at least two of the F -ratios for IPD in table 16.7 are 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 16.7 are indeed larger than the corresponding entries in table 16.4. The $\hat{\rho}_i$ estimates indicate the presence of a second-order autoregressive process in the error terms of the OLS regressions for UR.

16.3.3 Ex Ante Errors in Individual Forecasts

The actual values employed in the previous sections include all of the non-conceptual (prebenchmark) revisions in the data. These revisions presumably

Table 16.7 Regressions of Actual on Predicted Values and Tests of Bias, 18 Series of Group Mean Forecasts, GLS Estimates, 1968-79

Quarter Predicted	Regression Estimates		F-ratio for $\alpha = 0, \beta = 1$		t-Test for $\alpha = 0, \beta = 1$		Estimated Autoregression Coefficients			Squared Correlation (r^2) (9)	Durbin-Watson (10)
	a (1)	b (2)	(3)	(4)	(5)	(6)	(7)	(8)			
<i>Gross national product (GNP)</i>											
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°
<i>Implicit price deflator (IPD)</i>											
Q1	.60	.75	2.56 [§]	1.49	-.84	-.47*				.10	2.06°
Q2	1.15	.36	2.95 [§]	2.14*	-1.58	-.48*	-.07			.06	2.13°
Q3	1.03	.49	2.27	1.64	-1.04	-.59	-.12	.03		.08	2.12°
<i>Real gross national product (RGNP)</i>											
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°
<i>Unemployment rate (UR)</i>											
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*
<i>Consumer expenditures for durable goods (CEDG)</i>											
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°
<i>Change in business inventories (CBI)</i>											
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 sec. 16.3.2 on the model and computer procedure used. On the meaning of the measures and symbols, see notes to table 16.4.

bring the data closer to the "true" values that one would like to have predicted. (Indeed, some forecasters state for the record that they should be tested for their ability to predict revised, not preliminary, values.) But it is important to recognize that such data, and hence the estimates derived from them, are all *ex post* in nature. The residual errors from our regressions could not have been *known* to the forecasters on a 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.

The following tests allow for this problem by using series of past errors defined as in section 16.1.2 and 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 all systematic elements in it. It must be noted, however, 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 the creation of a comprehensive computer file of successive vintages of the data covered.

Drawing upon that record, the *ex ante* forecast error series for each individual and target variable are compiled. The autocorrelation functions for these errors are estimated as

$$(17) \quad \hat{\rho}_k = \frac{\sum_{t=1}^{T-k} \varepsilon_{t+j} \varepsilon_{t+j-k}}{\sum_{t=1}^T \varepsilon_{t+j}^2}, \quad j=0, \dots, 3; \\ k=j+1, \dots, j+6.$$

Here ε_{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 coefficient for the lag 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 $t + j$. 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; to $\hat{\rho}_2$ for Q2 and Q3; and to $\hat{\rho}_3$ for Q3.

The autocorrelation coefficients $\hat{\rho}_k$ 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 16.6. The test statistics Q_j are then calculated by (15) with $6-j$ df. The results are summarized in table 16.8.

The critical 10% significance level is widely used in practice as a cutoff for the Q -test, and on this criterion the joint hypothesis that all of the examined autocorrelation coefficients are 0 cannot be rejected for more than half of the

Table 16.8

Chi-Squared Tests of Autocorrelations of Errors in Selected Forecast Sets, Six Variables, 1968-79

Quarter Predicted	No. of Forecast Series ^a (1)	% of Forecasts with Q -Statistics ^b Significant at the Level of			% of Forecasts with $\hat{\rho}_k > 2/\sqrt{n}$ (5)
		1% (2)	5% (3)	10% (4)	
<i>Gross national product (GNP)</i>					
Q0	19	21.0	31.6	47.4	52.6
Q1	19	15.8	36.8	57.9	36.8
Q2	19	0	5.3	26.3	0
Q3	18	11.1	11.1	16.7	11.1
<i>Implicit price deflator (IPD)</i>					
Q0	20	45.0	60.0	60.0	60.0
Q1	20	20.0	45.0	80.0	60.0
Q2	20	30.0	50.0	50.0	50.0
Q3	20	45.0	60.0	75.0	85.0
<i>GNP in constant dollars (RGNP)</i>					
Q0	19	10.5	31.6	31.6	15.8
Q1	19	0	15.8	15.8	10.5
Q2	19	0	5.3	10.5	5.3
Q3	18	0	33.3	50.0	5.6
<i>Unemployment rate (UR)</i>					
Q0	18	22.2	27.8	27.8	22.2
Q1	18	22.2	33.3	38.9	33.3
Q2	18	33.3	44.4	44.4	55.6
Q3	17	23.5	52.9	58.8	47.1
<i>Consumer expenditures for durable goods (CEDG)</i>					
Q1	18	0	16.7	33.3	11.1
Q2	18	0	11.1	16.7	0
Q3	17	0	0	5.9	0
<i>Change in business inventories (CBI)</i>					
Q0	20	15.0	35.0	45.0	60.0
Q1	20	10.0	45.0	45.0	35.0
Q2	20	25.0	60.0	65.0	80.0
Q3	20	15.0	40.0	45.0	20.0
<i>Summary</i>					
Variable:					
GNP	75	12.0	21.3	37.3	25.3
IPD	80	35.0	53.8	66.2	63.8
RGNP	75	2.7	21.3	26.7	9.3
UR	71	25.4	39.4	42.3	39.4
CEDG	71	5.6	16.9	23.9	9.9
CBI	80	16.2	45.0	50.0	48.8

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

^aIncludes forecasts of those individuals who participated in more than 12 consecutive surveys. The few observations available for Q4 are excluded.

^bRefers to the Box-Pierce statistics as defined in eq. (15), with $6-j$ df. See text.

^cCovers all forecasts used in this table, summarized across the four target quarters, Q0-Q3.

error series covered. The best results are obtained for the RGNP and CEDG forecasts (see cols. 2–4 and the summary in table 16.8). In contrast, two thirds of the series for IPD, half or more of those for CBI, and the long forecasts of UR have Q -statistics that are significant. Thus in many cases forecasters appear to have failed to treat their own past errors efficiently as data to learn from. (One can think of various possible reasons for this: inconsistent or deficient information, models, and judgments, surprisingly large and frequent disturbances.)

It should be noted that these chi-squared tests are neither strong nor direct. (For example, a value of Q below the 10% level indicates a probability of less than 90% that the hypothesis that the errors are not white noise is true. For more detail and examples, see Pindyck and Rubinfeld 1981, pp. 549–50.) An additional test is performed by inspecting all individual $\hat{\rho}_k$ coefficients to see how many of them fall outside the range of 2 standard deviations from 0. The results, listed in the last column of table 16.8, generally agree well with our earlier conclusions.

16.3.4 Ex Ante Errors in Group Mean Forecasts

Table 16.9 presents sample estimates of the autocorrelation functions (eq. [17]) 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 .154. Of the 108 entries in columns 1–6 of the table, 82 are smaller than .154 in absolute value; 22 fall between .154 and .301; and only four exceed .301, that is, are outside the range of ± 2 standard deviations from the mean 0. Inflation forecasts account for eight of the observations in the second and all four observations in the third group.

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 (col. 7). In several cases, the calculated Q 's decline between Q_0 and Q_3 , 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 the autocorrelations of higher order deserve much attention, but it certainly cannot be assumed that all of them are 0. (In an earlier study based on ex post errors in the group mean forecasts and using as many as 12 autocorrelation lags, some of the $\hat{\rho}_k$ coefficients for $k > 6$ were found to be large and significant [see Zarnowitz 1982a, 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.)

Table 16.9 Tests of Autocorrelation of Errors in 24 Series of Group Mean Forecasts, 1968–79

Quarter Predicted	Estimated Autocorrelation Coefficients ^a						Box-Pierce Statistic ^b (Q_j) (7)
	$\hat{\rho}_1$ (1)	$\hat{\rho}_2$ (2)	$\hat{\rho}_3$ (3)	$\hat{\rho}_4$ (4)	$\hat{\rho}_5$ (5)	$\hat{\rho}_6$ (6)	
<i>GNP in current dollars (GNP)</i>							
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
<i>Implicit price deflator (IPD)</i>							
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 [#]
<i>GNP in constant dollars (RGNP)</i>							
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
<i>Unemployment rate (UR)</i>							
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
<i>Consumer expenditures for durable goods (CEDG)</i>							
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
<i>Change in business inventories (CBI)</i>							
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

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

^aFor level errors in UR and CBI and for percentage change errors in the other variables. All measures refer to the means of predictions by those individuals who participated in at least 12 surveys. See eq. (17) and text.

^bSee eq. (15) and text. # means significant at the 5% level, §, at the 10% level.

16.5 Summary and Conclusions

16.5.1 Main Results

1. The hypothesis of unbiasedness is rejected at the 5% significance level for 362 of the 2,350 forecast series examined. Nearly half of these rejections refer to the inflation forecasts, where they account for 44.3% of the tests. For the other five variables, the percentage of rejections is 9.6. It is evident that the forecasts of inflation show much more bias than the others.

2. The contrast extends to each target quarter but is particularly sharp for the longer horizons, as illustrated by the following percentages rejected:

	<i>Q0</i>	<i>Q1</i>	<i>Q2</i>	<i>Q3</i>	<i>Q4</i>
IPD forecasts	25	44	53	59	54
Other forecasts	12	7	8	9	9

3. About 4 out of 10 of the respondents participated in fewer than 20 of the surveys covered, producing relatively short forecast series for which the power of the tests is low. Higher significance levels and confidence region analysis are used to cope with this small-sample problem. The previously derived broad conclusions on the incidence of bias are not altered.

4. Inflation has been underestimated most of the time during the 1970s in both the short and, particularly, longer forecasts. Mainly for this reason, the mean errors are predominantly negative for all nominal variables covered. They tend to be positive for the predictions of rates of change in real GNP.

5. Because in each unit period there is only one actual value but many individual predictions for any target variable, pooled regressions yield biased estimates of the parameters. They result in unreliable tests that greatly overstate the extent of bias in the forecasts.

6. For inflation and unemployment, most of the regressions of actual on predicted values have serially correlated residual error terms, and the frequencies of this result increase with the predictive horizon (distance from the target quarter). For the other variables, only minorities of the tests indicate the presence of such dependencies. The latter increase the risk of falsely rejecting the null hypothesis of unbiasedness. However, cases in which separate tests indicate both bias and autocorrelated residuals appear to be relatively infrequent, except for inflation.

7. For the series of group mean forecasts, the tests do not reject $H_0: (\alpha, \beta) = (0, 1)$ for GNP (Q0–Q4), RGNP (Q2, Q3), UR (Q1, Q2), and CEDG (Q1–Q4). Elsewhere there are evidences of bias, weak in some cases and strong in others, such as the shortest forecasts for RGNP and CEDG, intermediate forecasts for CBI, and all forecasts (but particularly the longer ones) for IPD. These results are from tests based on OLS regression estimates; tests that use GLS estimates produced fewer and weaker rejections of the unbiasedness hypothesis.

8. 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, where such errors are measured strictly in the form in which they could have been known at the time of the forecast. Tests of autocorrelation among errors so measured are again unfavorable for most individual forecasts of IPD and many forecasts of CBI and UR and considerably more favorable for the other variables. There is no evidence of a common dependence of these results on the forecasting span.

9. The corresponding tests for the group mean forecasts confirm the negative verdict on the inflation expectations, but they find little evidence of serial correlation among the *ex ante* errors of the other forecasts for lags of up to six quarters.

16.5.2 Related Findings

Recent work using survey data has been preoccupied with tests for the rationality of inflation expectations. The most frequently used series consists of one-step-ahead group forecasts from the semiannual Livingston surveys. The balance of the evidence is that these data reject the hypothesis that $(\alpha, \beta) = (0, 1)$. This conclusion is consistent with the results of Pesando (1975), Carlson (1977), Wachtel (1977), Moore (1977c), Pearce (1979), Figlewski and Wachtel (1981), and Gramlich (1983). Tests that do not reject the hypothesis were reported by Mullineaux (1978, 1980a).

Another set of data much examined in this context comes from the surveys of consumer attitudes of the University of Michigan Institute for Social Research (ISR; Juster 1979; Huizinga 1980; Curtin 1982). The questions on "prices of things you buy," once concerned solely with the *direction* of the expected changes, now also seek to classify the rates of change into a few broad *size* categories. The creation of a group forecast series from such data requires a rather elaborate *ex post* procedure of quantifying qualitative responses. Some of the studies find that the unbiasedness hypothesis cannot be rejected for the ISR data; others merely find that it is "not so decisively rejected" as the inflation forecasts by other groups (Gramlich 1983, p. 163). In all, there is not much convincing evidence here in favor of the REH. Claims that households predict inflation "more rationally" than economic and business forecasters receive some support, but they are based on comparisons that seem strained and inconclusive.

The plant and equipment surveys of the Commerce Department's Bureau of Economic Analysis yield forecasts by businessmen of price changes for goods and services and capital goods purchased; these data fail to pass the test for unbiasedness decisively (de Leeuw and McKelvey 1981; Gramlich 1983).

Data from the European and Japanese surveys, mainly of consumers, require quantification of categorical responses. The results of the tests are here mixed but in large measure negative (Aiginger 1981; Papadia 1982; Visco 1984).

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.

McNees (1978) applied tests of unbiasedness to forecasts from three well-known econometric service bureaus, Chase, Data Resources, Inc., and Wharton. The periods covered are short—5.5 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 accepted for predictions over more than one quarter when GLS is used. The results for UR are mixed, with indications of bias in the longer but not the short forecasts.

According to a study by Hafer and Hein (1985), the average ASA-NBER forecasts of inflation were more accurate in 1975–79 and 1980–84 than in 1970–74. In the last two subperiods and in the full period 1970–84, they outperformed the inflation forecasts from a time-series (ARIMA) model and an interest-rate model (similar to Fama and Gibbons 1982). The first half of the 1970s was clearly among the most trying times for the forecasters generally (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.