Diversity, Uncertainty, and Accuracy of Inflation Forecasts

Stephen K. McNees with the assistance of Lauren K. Fine

Vice President and Economist and Research Assistant, respectively, Federal Reserve Bank of Boston. *The author gratefully acknowledges helpful comments from J. Joseph Beaulieu, Dean Croushore, Ray Fair, and his colleagues at the Federal Reserve Bank of Boston, in particular Stephen Blough.*

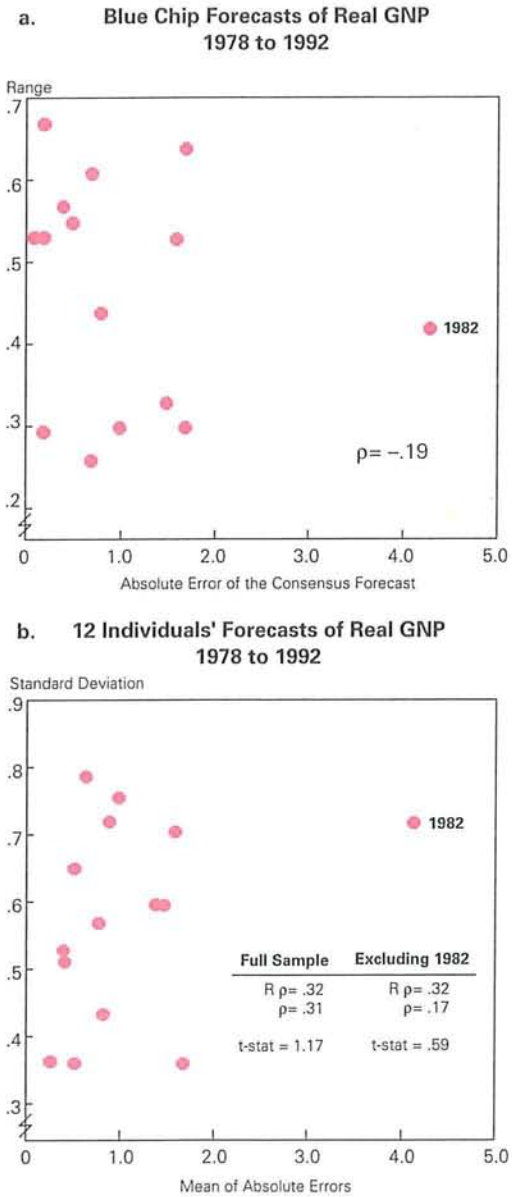
Early each year, numerous surveys of economic forecasts are published. This year, not only did the surveys in *Business Week*, *The Wall Street Journal*, and *Blue Chip* all show exactly the same central tendency (the consensus forecast), but the dispersion among the forecasts was extraordinarily small. Such conformity appears to lend reliability to the forecasts; after all, if no one expects an outcome much different, surely the consensus view is relatively certain.

Unfortunately, this may be an instance where common sense is misleading. Figure 1a presents the range of the Blue Chip real GNP forecasts made each October (the month that Robert Eggert, the collector of these forecasts, emphasizes in his retrospective analyses) since 1977, along with the absolute value of the error of the Blue Chip consensus forecast. The figure clearly illustrates and a simple correlation confirms that little relation exists between the range of Blue Chip GNP forecasts and their eventual accuracy; indeed, the simple correlation between the two is negative 0.19. Comparable correlations for other variables (nominal GNP, the implicit GNP deflator, the unemployment rate, nonresidential fixed investment, and housing starts) also exhibit no significant relationship between the dispersion of the individual forecasts and the accuracy of the Blue Chip consensus forecast—the highest positive correlation is only 0.19 for nominal GNP and becomes negative 0.31 when the outlier forecast of 1982 made in October 1981 is dropped.

This lack of correlation should come as no surprise for at least two reasons. First, the Blue Chip survey group is not a fixed set; forecasters come and go and participation rates of forecasters vary over time. If the new entrants differ from the dropouts, the characteristics of the group would change. Figure 1b presents roughly the same information as Figure 1a for a subsample, 12 forecasters who forecasted all six variables every year—11 Blue Chip participants and one prominent forecasting group that does not regularly participate in the Blue Chip survey. In this case, the dispersion of the individual forecasts is measured by their

Figure 1

Dispersion and Accuracy



standard deviation and their accuracy by the mean of their absolute errors. Even though the simple correlations are higher for this subsample of regular participants, the highest correlation among the six variables, 0.33 for nominal GNP, is still not close to statistical significance.

The lack of a significant correlation also reflects a conceptual problem. The foregoing analysis, like much of the literature, uses the dispersion of individual point estimate forecasts as the measure of forecast uncertainty. In fact, the dispersion of individual forecasters' point estimates may measure conformity but not necessarily uncertainty. (See Zarnowitz and Lambros 1987, especially their Figure 1.) It is entirely possible for all forecasters to share a common point estimate yet also acknowledge that they are highly uncertain that their forecast will prove to be reliable; this may well describe stock market and exchange rate forecasts. At the other extreme, one can easily imagine two dogged forecasters each expressing great confidence in the accuracy of their highly divergent point estimate forecasts. In short, uncertainty and the conformity of point estimate forecasts are logically distinct concepts, so that the type of evidence discussed so far does not directly address the relationship between forecast uncertainty and the accuracy of a point estimate forecast.

I. Why Forecast Uncertainty Is Important

Economic forecasts are nearly always expressed as a single number or "point estimate." A point estimate constitutes a very limited description of the entire array of possible (or even plausible) future outcomes. To a statistician, a distribution of the probabilities of alternative outcomes is poorly characterized by its central tendency alone.

Both forecasters and forecast users have suffered from the concentration on point estimates and the lack of attention to a fuller array of expected outcomes. The problems with point estimates are both substantive and practical. Forecasters are asked to predict a wide range of data, which range from highly volatile (random, or "unpredictable") to stable, "well-behaved" series. For example, the semiannual *Wall Street Journal* forecast survey asks for estimates of both exchange and interest rates and inflation and unemployment rates. Under the Efficient Markets Hypothesis, changes in the former are random or near-random series, whereas the latter show great persistence and can, in fact, be predicted much more accurately than by naive rules of thumb (McNees 1992, Table 4). My exchange rate forecast may be as good as any other yet no better than a coin flip; my inflation rate forecast need not be the best available to dominate simple rules of thumb.

Failure to communicate these differences in "forecastability" or, more precisely, failure to communicate some information about the expected uncertainty or reliability of a forecast not only undermines the utility of the forecast but also risks providing a distorted picture of the forecast's reliability to the forecast user. If you think that my stock market forecasts will be as good as my inflation rate forecasts, you will be likely to overestimate my ability to forecast stock prices. When you eventually learn of the limitations of my stock price forecasts, you will undervalue my inflation rate forecasts unless you realize that my inflation forecasts have relatively high reliability.

Forecasts are commonly used for risk management or contingency planning. No matter how accurate, a point estimate alone cannot convey how to incorporate uncertainty in decisionmaking.

II. Measures of Uncertainty

Forecasts inherently are linked to uncertainty; the reason to forecast is that the future is uncertain and unknowable. To say that the future is unpredictable is a non sequitur; the fact is that all future events are unknown, so that any discussion of the future—any plan—is essentially a forecast, subject to error.

Although the meaning of uncertainty is fairly clear, its measurement is problematic. According to my dictionary, "Uncertainty may range from a falling short of certainty to an almost complete lack of definite knowledge, especially about a [future] outcome or result." This observation not only illustrates the inexorable link between forecasts and uncertainty but suggests that the concept of uncertainty may not have a unique empirical counterpart; that is, it covers a range of different levels of uncertainty or a family of different orders of uncertainty, as will be discussed further below.

In practice, uncertainty has been measured in several different ways. In the academic literature, the most common method is to examine the variability of the residuals of a stochastic econometric model. This method has one obvious problem and one more subtle problem. The obvious problem is that one must assume that the specified model is the one actually used to form expectations. This assumption raises a host of standard issues: Does the model incorporate all relevant informational variables? Are they weighted appropriately? Do all individuals use the same model to formulate their expectations? More basically, even if we were prepared to assume that a particular model is the correct description of the universal expectations formation mechanism, one could question whether the historical residuals of that model are good measures of real-time, post-sample uncertainty. The problem is that the future residuals are a complicated combination of the effects of (largely) unanticipated events, partially anticipated events, and any "white noise," irreducible stochastic error.

A wholly unanticipated event clearly would affect the residual but, by definition, would not be a source of ex ante uncertainty—what you cannot conceive of cannot worry you. Real world examples of largely unanticipated events that probably had nontrivial impacts on macroeconomic forecasts in the past 25 years include the outbreak of wars (the Yom Kippur war, the revolution in Iran, the Iraqi invasion of Kuwait) as well as surprise shifts in economic policy (the 1971 Nixon wage-price freeze, the 1979 change in Federal Reserve operating procedures, the 1980 imposition of credit controls). The list of partially anticipated events is much longer; virtually any change in fiscal policy has been anticipated with varying degrees of uncertainty over time as proposals for legislation pass from the executive branch through both houses of Congress, a reconciliation process, and a Presidential veto or signature.

Failure to communicate information about the expected uncertainty of a forecast undermines its utility and risks providing a distorted picture of its reliability to the forecast user.

A particularly clear example is the Reagan personal income tax cuts. Reagan proposed the cuts in his 1980 campaign for President, raising the possibility of substantial tax rate reductions long before they went into effect. Over time, economic analysts' estimates of uncertainty with regard to tax rates diminished as the chances of Reagan's election and the prospects for congressional approval improved. While many models include present and future tax rates, none are well equipped to measure the political-economic evolution of uncertainty about future tax rates over time. Fiscal policy is by no means the only example of partially anticipated events. The onset and resolution of major strikes, monetary policy changes, and the relaxation of temporary programs like wage and price controls or credit controls all fall into the category in which the uncertainty of a macroeconomic event evolves gradually toward zero over time. This learning process seems to be the essence of what we mean by uncertainty with regard to future events. It is difficult to see how partially anticipated events such as these, which were without close precedent in the sample period to which the models were fit, can be measured by the historical residuals of any macroeconometric model.

In light of the difficulty of modeling uncertainty, a plausible and easily obtainable measure of uncertainty has been the dispersion of individuals' forecasts. This procedure presumes that when different individuals have unusually great dispersion among their point estimate forecasts, then uncertainty is high. It should be clear that this measure is, at best, only a crude proxy for uncertainty. As previously noted, no logically necessary connection exists between a forecaster's degree of uncertainty and the degree of uniformity of the point estimates of different forecasters. It is perfectly possible for each forecaster to be wholly confident in his forecast yet for all forecasters to hold widely varying views. Similarly, all forecasters could easily agree that an exact number is the best single point estimate forecast but also that the degree of certainty of that forecast is extremely low—total conformity and high uncertainty.

III. Description of the Data

To our knowledge, the only systematic collection of real time or ex ante estimates of the uncertainty of macroeconomic forecasts is the Survey of Professional Forecasters, originated by Zarnowitz in 1968 (see Zarnowitz 1969) and maintained by the Federal Reserve Bank of Philadelphia since late 1990. (See Croushore 1993 for an excellent description of this data set.)

In addition to a point estimate forecast, each survey respondent is asked to provide two types of probability distributions: (1) the probability that real GDP will decline in the current quarter or in any of the next four quarters (these data are analyzed in Braun and Yaniv 1992), and (2) probability distributions of the expected year-over-year percent change in the GDP deflator (here called inflation uncertainty) and in nominal GNP (until 1981:III, when real GNP replaced nominal). Specifically, each respondent distributed probability across 15 intervals 1.0 percentage point in width.[1,2] This study analyzes the second type of probability distributions, those for the annual percent changes in the GDP deflator (from 1968 to 1993) and real GNP (1982 to 1993).

Because these distributions pertain to year-over-year realizations and the survey is conducted quarterly, four consecutive surveys provide four estimates of the same realization taken from four different forecast horizons. From 1968 through 1981, the first estimate for each year—the longest forecast horizon—was the first survey taken early in that year, shortly after the actual data for the final quarter of the preceding or base year became available. (Since 1981:III, the forecast horizon has been extended to two years overall, so that each year now is surveyed eight times.) The three subsequent surveys were taken well within the year and thus combine a forecast of the remainder of the year with partial, actual data for part of the year. Clearly, the amount of uncertainty varies considerably across horizons that embody varying amounts of actual data. This study focuses exclusively on the probability distribution with the four-quarter-ahead forecast horizon, the one that is not intermingled with partial, actual data. Thus, the maximum number of observations for an individual forecaster is 25, corresponding to the years 1969 through 1993.

Several authors have examined the probability distributions of the forecasts, including Lahiri and Teigland's 1987 paper concluding that the means of the distributions were not normal, and Lahiri, Teigland, and Zaporowski's 1988 study concluding that real interest rates decline when inflation uncertainty increases. The previous study most comparable to this one is by Zarnowitz and Lambros (1987).

Previous studies based on this data set have reported results for aggregations of the individual responses, such as the mean value in each survey. This choice was dictated in part by the paucity of data for individual forecasters available at the time. Aggre-

[1] In 1981:III, a redesign of the survey reduced the number of intervals to six and increased their width to 2.0 percentage points. In 1992:I, the width of the intervals was returned to 1.0 percentage points and their number increased to 10.

[2] Each survey contains extreme intervals that extend to plus or minus infinity; in order to calculate means and standard deviations, it was assumed that the width of these extreme intervals was the same as the intermediate intervals—that is, usually 1.0, but where appropriate 2.0, percentage points. This truncation of the extreme intervals does not appear to cause a distortion; the actual outcome never fell outside the range defined by these truncated extreme intervals.

gate data inevitably reflect the erratic response rate from the participating forecasters. When the survey started in 1968, more than 50 individual forecasters participated; the number of participants had declined to fewer than 20 by 1988; then, when the Philadelphia Fed revived the survey, the number of participants rose above 30 (Croushore 1993, p. 4). Over time, more than 300 different forecasters have participated in at least one of the surveys; however, about half have participated in only 10 or fewer surveys.

It is not clear whether the low and variable participation rates of these forecasters reflect a lack of interest in macroeconomic forecasting or these individuals simply found no incentive to share their forecasts with the survey. (All respondents are granted strict confidentiality.) Erratic participation would be a particular problem if it were systematically related to, for example, the perceived uncertainty of the forecast period. In any event, it seemed wise to exclude the infrequent participants, following Zarnowitz and Lambros (1987, pp. 602–3), who eliminated "occasional" forecasters who had participated in fewer than 24 percent of the 51 surveys through 1981:II in order to reduce "the variation in coverage over time." The criterion applied here is participation in at least 10 of the 25 possible forecasts.

The focus of this study is individual forecasters' inflation uncertainty measured by the probability distributions of the year-over-year GNP deflator forecasts made early each year, as soon as the previous year's fourth-quarter data became available. One hundred and fifty different respondents provided 905 such forecasts; excluding individuals who provided fewer than 10 left a sample of 33 forecasters and 394 probability distributions, which are described and assessed in the next section.

IV. Reliability of the Uncertainty Estimates

It was argued above that a realistic estimate of uncertainty is inherently individualistic or "subjective"; that neither a backward-looking time series model nor a conditional, structural model is likely to capture the major sources of uncertainty that arise in practice. Most uncertainty seems to arise from extra-model sources, such as a lack of information with respect to the future values of exogenous variables, or from the conviction (sometimes even the knowledge) that a change in the structure has occurred, so that the model's description of the past is unlikely to prevail in the future. For this reason, this study concentrates on the individual responses to surveys that simply ask professional forecasters to estimate, as best they can, the distribution of alternative future outcomes.

Although this emphasis on individuals rather than statistical models differs from that followed by most economists and statisticians (see Chatfield 1993), it is entirely consistent with the perspective of most psychologists. Hogarth (1980, pp. 11–12), for example, emphasizes that a probabilistic statement expresses our degree of knowledge and is not "a property of events in the environment." Precisely

Most of these professional forecasters clearly possess an ability to anticipate the uncertainty of their forecasts.

because uncertainty and probabilistic statements designed to reflect uncertainty are individualistic or subjective, it seems critically important to examine whether these data have any "validity"—whether they have intrinsic coherence and conform with reality. Because the "true" probability distribution of alternative outcomes cannot be observed in nonexperimental social sciences, no airtight validity check is available. Instead, we employ a battery of checks.

At the crudest level, the data were screened to ensure that the probabilities provided summed to 100 percent. Five exceptions that could not be attributed to rounding were found; the two instances found in the subsample of 33 forecasters examined here were discarded.

At a slightly more substantive level, uncertainty should show some variation over time. It would be quite dubious to discover that the estimated level of uncertainty was the same in the mid 1980s as it was in the aftermath of Nixon's August 1971 New Economic Policy, the first OPEC oil embargo and price shock, the Fed's October 1979 change in operating procedures, Carter's credit controls, or Reagan's fiscal revolution.

In addition, even though uncertainty estimates are individual, they would be of little use if no conformity among individuals occurred over time. While uncertainty estimates need not be identical among individuals, they must move broadly together in order to render meaning to statements like "These are highly uncertain times." (I do not recall anyone ever saying, "These are highly certain times.")

Table 1 illustrates both the variety and the conformity of the individual forecasters' inflation uncertainty estimates. On the one hand, conformity is far from total: In only two years (1985 and 1993) did all of these respondents record below-average levels of uncertainty. In 10 years, at least one forecaster attached the maximum uncertainty estimate to his forecast while another attached the minimum. On the other hand, seven of the 12 respondents for 1982 attached more uncertainty to inflation in that year than in any other year. Five of 19 found uncertainty highest in 1975. Ten of the 17 respondents in 1986 attached the lowest uncertainty to estimates in that year. All respondents felt uncertainty was below normal in 1985, with six ranking it lowest of the years in which they responded. The standard deviation for 10 of the respondents for 1986 was 0—or, in other words, 100 percent of the probability was assigned to one interval. Thus, while evidence of variability across individuals is ample, enough conformity is also to present suggest that generalizations about the prevailing degree of uncertainty are usually warranted. The mean of all respondents each year, shown in the last row of the table, reached a high in 1980 and a low in 1993.

Table 1
Individual Forecasters' Inflation Uncertainty Estimates
33 Forecasters, 1969 to 1993

		High		Low	
	$\overline{SD}$	SD_H	Year(s)	SD_L	Year(s)
	.41	.60	1982 and 1984	0	1986
	.57	.92	1989	.43	1976
	.62	.79	1970	.46	1972
	.62	1.58	1989	0	1986
	.62	1.00	1986	.45	1970 and 1973
	.62	1.01	1973	0	1986
	.62	1.00	1975	.32	1973
	.64	1.15	1975	0	1986
	.66	1.73	1975	0	1982 and 1985 to 1988
	.68	1.28	1977	0	1985 to 1986
	.69	1.37	1972	0	1985 to 1986
	.71	1.93	1983	0	1986
	.72	1.59	1982	0	1985 to 1986
	.74	1.31	1982	0	1984 and 1986
	.74	1.00	1971	.60	1993
	.78	2.67	1979	0	1988
	.81	1.57	1983	.59	1970 and 1974
	.82	1.40	1982	.30	1969
	.82	1.19	1975	.65	1980
	.87	1.39	1971	.54	1980
	.87	1.68	1980	.40	1969
	.88	1.31	1982	.70	1973
	.93	1.24	1971	.73	1970
	.96	1.95	1990	.40	1985
	1.01	1.61	1982	.64	1969
	1.03	2.39	1988	.30	1983 to 1984
	1.06	1.84	1979	.55	1972
	1.11	2.22	1982	.67	1976
	1.14	1.51	1981	.65	1973
	1.20	1.99	1975	.48	1973
	1.26	2.18	1984	0	1985
	1.51	2.09	1974	.98	1971
	1.52	2.57	1983	.30	1969
Mean	1.23	1.94	1980	.62	1993

Though illustrative, these facts do not take advantage of the fact that specific, quantitative confidence limits are available. For any given level of confidence, a binomial test can be used to determine whether each forecaster's estimated confidence interval was statistically significantly different from the actual outcome. For any given confidence limit, each forecaster has a corresponding forecast interval, within which the actual value either will (a hit) or will not (a miss) fall, a binomial outcome. If a forecaster has estimated or calibrated his uncertainty accurately (that is, he is neither overconfident nor overcautious), the actual value should fall within the forecaster's 90 percent forecast interval 90 percent of the time *for a large sample of forecasts*. Of course, with a limited sample, the observed percentage of hits could differ from 90 percent by statistical chance rather than because the forecaster is systematically overconfident or overcautious. The binomial distribution can be used to test whether deviations from the predicted 90 percent occur by chance or not. Specifically, let

N = the number of observations,
M = the number of actual outcomes within the forecast interval, or "hits," and
p = the theoretical probability of a hit.

Hence, the binomial distribution B(N,M,p) or

$$Prob\ (hits = m) = \binom{N}{M} p^M (1-p)^{N-M}$$

gives the probability of observing exactly M hits out of N trials when the true probability of a hit is p.

For example, suppose a forecaster has made N forecasts and the actual value has fallen within the specified 90 percent interval M times. The probability of observing M *or fewer* hits if the true probability of a hit is 0.9 is:

$$P^L(0.9) = \sum_{m=0}^{M} B(N,m,0.9).$$

If this quantity is small, it is unlikely that the forecaster has had so few hits by chance. Formally, if, for example, P^L is less than 0.05, we can reject at the 5 percent level the hypothesis that the forecaster's specified 90 percent confidence intervals are truly 90 percent confidence intervals, in favor of the alternative that the true probability of the interval covering the actual value is less than 90 percent. The forecaster is *significantly overconfident*—the actual value falls inside the specified 90 percent confidence intervals too rarely.

Conversely, the probability of observing M *or more* hits if the true probability of a hit is 0.9 is:

$$P^H(0.9) = \sum_{m=M}^{N} B(N,m,0.9).$$

If this quantity is small, it is unlikely that the forecaster has had so *many* hits by chance. If P^H is less than 0.05, we can again reject the hypothesis that the specified 90 percent confidence intervals are truly 90 percent confidence intervals, but this time in favor of the alternative that the true probability is *greater* than 90 percent. The forecaster is *significantly overcautious*.

Note that, since the total probability of all possible numbers of hits is 1,

$$P^H(p) = 1 - P^L(p) + B(N,M,p).$$

The above discussion is based on one-tailed tests. In this case, a two-tailed test seems appropriate, in that we wish to penalize not only the overconfident forecaster who tells us the risks are smaller than they actually are but also the overcautious forecaster who portrays the risks as greater than we need to fear. For a two-tailed test, at the 5 percent level, if either P^L or P^H is less than 0.025, the null is rejected in favor of the two-sided alternative.

Table 2 gives the results of the binomial test for the accuracy of the individual forecasters' inflation uncertainty forecasts. The results show conclusively that these uncertainty estimates are "valid," that is, that they generally conform quite closely to reality. For example, at the 90 percent confidence limits, for 88 percent (29 of 33) of the forecasters one can reject the hypothesis that the forecaster was either overconfident or overcautious, at a 95 percent level of confidence. In addition, at the 50 percent confidence limits, 94 percent of the forecasters were neither overconfident nor overcautious. Only at the 100 percent confidence limit was evidence of overconfidence widespread—about 70 percent of the forecasters experienced at least one actual outcome outside their entire distribution.

The bottom portion of Table 2 also notes that 30 percent (10 of 33) of the forecasters showed no evidence of either overconfidence or overcaution at any of these three confidence limits. Only two showed overconfidence at all three confidence limits. Had these outliers been typical, one could easily

Table 2
Calibration of Inflation Uncertainty Forecasts: Binomial Test Results
33 Forecasters, 1969 to 1993

	Overconfident	Neither	Overcautious
50% level	6%	94%	0
90% level	12%	88%	0
100% level	70%	30%	n.a.

Number of Forecasters	Description
19	Forecasters were neither overconfident nor overcautious at the 50 and 90 percent confidence levels but were overconfident at the 100 percent level.
10	Forecasters were neither overconfident nor overcautious at the 50, 90, and 100 percent confidence limits.
2	Forecasters were neither overconfident nor overcautious at the 50 percent limit but were overconfident at the 90 and 100 percent intervals.
2	Forecaster was overconfident at the 50, 90, and 100 percent confidence intervals.

n.a.—not applicable.

question whether this data set was worth investigating. The fact that only two of 33 forecasters did not provide realistic estimates of the reliability of their forecasts suggests that most of these professional forecasters do indeed possess some ability to anticipate the uncertainty of their forecasts.

An important exception to the general pattern of good calibration is, of course, the 100 percent confidence limits case where, for 70 percent (23 of 33) of the forecasters, the hypothesis that they are not overconfident was rejected. This fact may reflect a tendency to underestimate the (small) probability of highly unusual events. Even though we all "know" that "anything can happen," we often forget or ignore it in practice. This was certainly the case in 1973, a year when wage and price controls were relaxed, the first OPEC oil shock occurred, and a worldwide commodity price boom took place. Sixty-two percent (33 of 53) of the outcomes outside the 100 percent confidence interval took place in 1973 and 1974. Excluding these two years, 21 of the 33 forecasters showed no signs of overconfidence and the instances of overly optimistic 100 percent confidence were widely scattered over time.

It is tempting to think of 100 percent confidence limits as an unrepresentative, degenerate case. In fact, it serves as a useful reminder of the need to define uncertainty quite precisely. Possibly, individual forecasters could defend their overly optimistic 100 percent confidence limits ex post on the grounds that they had either explicitly or implicitly assumed no catastrophes. If explicitly asked, many might well have freely acknowledged that their uncertainty estimates were based on several implicit assumptions, such as no war in the Middle East, no dismantling of wage and price controls, no nuclear war, and the like. A skeptic could suggest that this amounts to a failure to understand what 100 percent confidence means; a more sympathetic and potentially more useful response is the recognition that there are differing, important levels of uncertainty. Even though it might be tedious to repeat before each forecast, all forecasts assume no nuclear war, no rapid global warming or Ice Age, no Black Plague epidemic, and so on ad infinitum. Yet some of these events do have a nonzero probability. Exactly where to draw the line between tediousness and rigorous precision—what is an appropriate, practical definition of uncertainty—is seldom discussed, though by no means obvious.

It is interesting to note that the tendency for most individuals to underestimate the 100 percent confidence limits does not appear in an analysis of the mean aggregate probability distribution of all respondents to each survey. The mean inflation uncertainty forecasts are well calibrated, neither overconfident nor overcautious, at the 50, 90, and 100 percent confidence limits. Even though most individual forecasters have been overconfident at the 100 percent level, at least one forecaster in each survey has assigned some probability to the interval in which the actual outcome fell. This is a clear illustration of how the aggregated, mean probability distribution can give a misleading picture of the underlying, constituent individual probability distributions.

V. Is Point Estimate Accuracy Related to Uncertainty?

Having established that most of the forecasters did in fact make plausible estimates of the uncertainty of their inflation forecasts, we return to the question that originally motivated this inquiry: Is uncertainty systematically related to the accuracy of point estimate forecasts? The relationship between inflation uncertainty (as measured by the standard deviation of the probability distribution of expected inflation) and accuracy (of the point estimate forecast) was examined by calculating both rank correlations and simple correlations over time between each forecaster's inflation uncertainty estimate and the accuracy of his point estimate forecast, as measured by either the absolute value forecast error or the squared value of the point estimate forecast error for the same year.

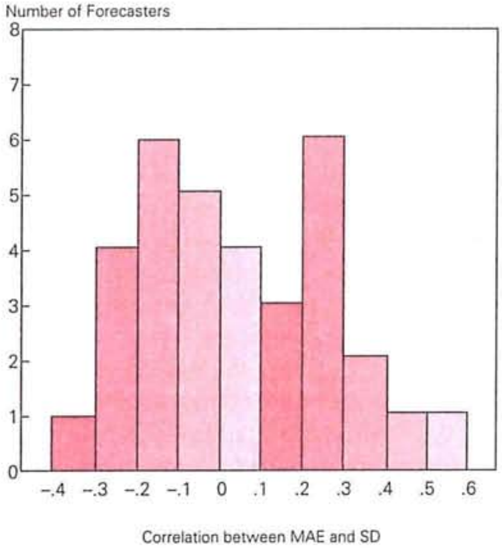
The results, summarized in Figure 2 and Table 3, show that little relationship exists between expected inflation uncertainty and the ex post accuracy of point estimate forecasts. The vast majority of these correlations were low and statistically insignificant. The more logical positive correlations far exceeded negative correlations for those few correlations that attained marginal significance. The closest resemblance to a positive relationship was for one forecaster whose rank correlation was 0.54, about 1.8 times its standard error, and whose simple correlation was 0.51 for absolute value of the forecast error and 0.40 for squared forecast error.

It is not entirely clear why uncertainty should be unrelated to forecast accuracy. One possible reconciliation of these apparently contradictory facts is that forecasters' point estimate forecasts represent the mode of their probability distribution and are fairly insensitive with respect to the width of the tails of the

Figure 2

Uncertainty and Accuracy of Inflation Forecasts

Frequency Distribution of Correlations



distribution. (In response to a survey conducted by the authors, two-thirds of the Survey of Professional Forecasters participants did describe their point estimate forecast as the *mode* of their probability distribution.) Forecasters have some ability to gauge when the tails of the distribution are fat and when they are thin, but this judgment has little impact on the mode or most likely single outcome. This would seem especially likely if the point estimate were determined independent of and logically prior to the entire distribution.

In any event, the importance of the results presented here is clearly limited by the small number of observations on which they are based. Further evidence may overturn them. Nevertheless, without contrary evidence, it would appear that estimating inflation uncertainty, even estimating it reliably, and selecting an accurate central tendency or point estimate forecast from that distribution are unrelated, disparate aspects of forecasting.

It would appear to be a mistake, however, to think that estimating uncertainty is unimportant simply because it is unrelated to the accuracy of the point estimate forecast. It seems more helpful to consider forecasting accuracy as having several distinct facets—including both point estimate accuracy and uncertainty reliability—which are not easily combined into one.

This conclusion is based on forecasts of the inflation rate, as measured by year-over-year percentage changes in the implicit GNP price deflator, the variable for which the most observations are available. Since 1981, the Survey of Professional Forecasters respondents have also provided estimates of probability ranges for real GNP. Eleven forecasters have provided eight or more real distributions analogous to the GNP deflator distributions discussed above.

As illustrated on Figure 3 and Table 4, this even more limited sample confirms the basic results derived from the inflation data: First, based on the binomial test, all 11 forecasters were well calibrated at the 50 percent confidence limit, and 10 of the 11 shared this characteristic at the 90 percent confidence limit. To the contrary, all 11 forecasters were overly confident at the 100 percent confidence limit. Once again, the aggregated, mean probability distribution was well calibrated at all levels, including the 100 percent confidence limit, reflecting the fact that at least one forecaster in each survey assigns some probability to the interval in which the actual outcome falls. As before, the general reliability of the estimated confidence intervals, except for the 100 percent level, provides justification for regarding the probability distributions as just as fundamentally important as the more widely publicized point estimates. Second, the overconfidence was heavily concentrated in a few years: Nearly one-half of the overconfident estimates of real GNP came in 1985, when the actual outcome fell outside the 100 percent confidence limits of eight of these 11 forecasters; virtually all of the remaining instances occurred in

Table 3

Correlations, Inflation Uncertainty, and Point Estimate Accuracy

33 Forecasters, 1969 to 1993

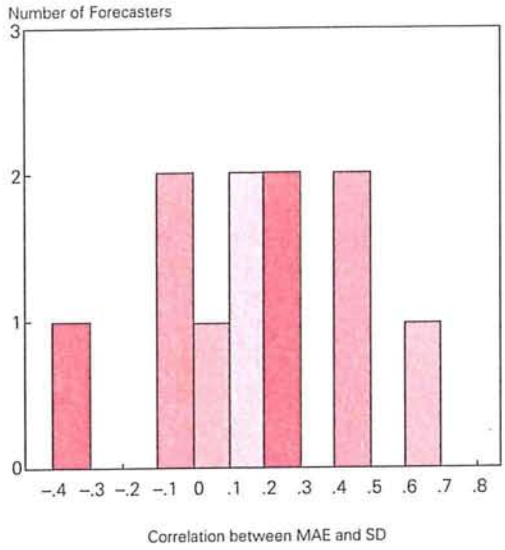
Level of Significance	$R\rho(U_t,\|e\|_t)$	$\rho(U_t,\|e\|_t)$	$\rho(U_t,e^2_t)$
"High" Positive[a]	2	1	0
"Moderate" Positive[b]	9	9	9
"Moderate" Negative[b]	6	6	3
No Statistical Significance	16	17	21

[a]Significant at the 5 or 10 percent level.
[b]Significant at the 20 or 50 percent level.

Figure 3

Uncertainty and Accuracy of Real GNP Forecasts

Frequency Distribution of Correlations



either 1982 or 1988. Finally, once again, with the exception of one forecaster for whom a statistically significant positive relationship was found, the amount of uncertainty was unrelated to the accuracy of the point estimate forecast, as measured by either its absolute or its squared error. Thus, the results for real GNP seem entirely consistent with the more extensive results for inflation.

VI. Are Accurate Point Estimate Forecasters Also Reliable Uncertainty Forecasters?

This paper opened with the observation that the probability distribution of alternative outcomes contains much more information than a point estimate forecast. Indeed, it is not even clear which, if any, of the standard measures of the central tendency of a probability distribution—the mean, the median, the mode, or some other measure—the point estimate forecast represents. Conceptually, forecast users could learn much more from the entire distribution than from a single point estimate. This conceptual advantage would be of little practical importance if estimated probability distributions (estimated uncertainty forecasts) were highly unreliable—if forecasters were systematically either wildly overconfident or overcautious. Yet the evidence presented in section IV above shows that, with minor exceptions, forecasters' uncertainty estimates are neither excessively bold nor excessively timid. Their uncertainty estimates were quite reliable, except perhaps for the 100 percent confidence intervals.

The absence of excessive confidence and caution is clearly a necessary condition for a good uncertainty estimate. Once this condition has been satisfied, the greater the confidence (the smaller the standard deviation of the probability distribution) that can be placed in a forecast distribution, the more helpful that forecast is to the ultimate forecast user. Thus, once the overconfident forecasters have been excluded, the smaller the standard deviation of the probability distribution, the better the forecaster's uncertainty estimate.

The previous section examined the relationship over time between individual forecasters' estimates of uncertainty and the accuracy of their point estimates. If all forecasters predicted all years, forecast users would prefer the forecaster whose standard deviations were smallest. Unfortunately, all forecasters in this data set include many "gap" years in which no uncertainty estimate is recorded. Some forecasters joined after the survey started, others dropped out, and many of the participants skipped a year occasionally. As we have seen, both forecast uncertainty and point estimate accuracy vary considerably from year to year. A simple unweighted standard deviation runs the risk of rewarding forecasters who participated in the relatively certain years and penalizing those who participated in the years of relatively high uncertainty. We have therefore weighted each forecaster's estimated standard deviation each year by

Table 4

Correlations, Real GNP Uncertainty, and Point Estimate Accuracy

11 Forecasters, 1982 to 1993

Level of Significance	$R\rho(U_t,\|e\|_t)$	$\rho(U_t,\|e\|_t)$	$\rho(U_t,e^2_t)$
"High" Positive[a]	1	1	1
"Moderate" Positive[b]	2	4	4
"Moderate" Negative[b]	0	1	1
No Statistical Significance	8	5	5

[a]Significant at the 5 or 10 percent level.
[b]Significant at the 20 or 50 percent level.

the average standard deviation of all forecasters who participated that year, in effect weighting each observation by its deviation from the mean for that year. The presence of data "gaps" also affects measures of point estimate accuracy. Thus, we also weight both the absolute error and the squared error by the average error of all forecasters who participated that year.

Once again, the results are so clear that they can easily be summarized; no relationship could be found across forecasters between the amount of uncertainty of their forecasts and the accuracy of their point estimates, as measured by either their mean absolute errors or their root mean squared errors. Both rank correlations and simple correlations are essentially zero, far from any significance in the statistical sense. Once again, point estimate accuracy and uncertainty accuracy are on skewed planes. One should not jump to the conclusion that uncertainty estimates are of no value simply because they do not predict point estimate accuracy. Indeed, one of the main premises of this inquiry is that the opposite conclusion would come closer to the truth; if a forecast can provide a reliable, well-calibrated expected distribution of outcomes, any measure of the relationship between its central tendency and the actual outcome may be of little interest.

VII. Summary and Conclusions

Uncertainty is a key concept in both economic theory and economic practice. Nonetheless, the premise of this article is that too little attention has been paid to defining and measuring the concept of uncertainty. This article has argued that many of the traditional measures of uncertainty are conceptually flawed and bear little empirical resemblance to actual uncertainty.

One of most fruitful empirical measures of macroeconomic uncertainty is the Survey of Professional Forecasters, started by Zarnowitz in 1968. Although the survey is conducted quarterly, it collects distributions of year-over-year changes and thus, for each forecast horizon, provides a single observation each year. The primary limitation of the survey is the uneven response rate; many forecasters participate only sporadically and even the regular participants skip occasionally. Their participation performance alone strongly suggests heterogeneity among the individual forecasters. The analysis in this study is confined to the individual forecasters with the highest participation rates.

Most forecasters had quite accurate estimates of the 50 and 90 percent confidence intervals of their inflation and real GNP uncertainty forecasts; few exhibited overconfidence and none showed overcautiousness. Virtually all of the individual forecasters were overly confident at the 100 percent level, a tendency not revealed in examining the mean probability distribution of all respondents to each survey. Somewhat surprisingly, an individual forecaster's uncertainty estimates are not highly correlated with the accuracy of his point estimates. This result emerges both for inflation forecasts and, with an even more limited sample, for real GNP forecasts. This result is consistent with the possibility that forecasters' point estimate forecasts are of the mode of the distribution, or some other measure of the central tendency not strongly related to its dispersion. In short, point estimate accuracy and uncertainty accuracy may well be two totally separate, disparate aspects of forecast accuracy.

Point estimate accuracy and uncertainty accuracy appear to exist as two totally separate, disparate aspects of forecast accuracy.

Even though both overconfidence and overcautiousness limit the usefulness of an estimate of uncertainty, among the estimates that do not exhibit overconfidence, smaller estimated uncertainty is preferable to larger estimated uncertainty. Once the overly confident forecasters are eliminated, the forecasters can be ranked by the size of their uncertainty estimates, and those with lesser uncertainty were found to provide more valuable information. These uncertainty rankings were not systematically related to the accuracy of point estimate forecasts ranked either by the absolute size of the error or by the squared error of the point estimate forecast.

The small sample on which these conclusions are based is an obvious limitation on their validity. Further evidence will be required to reach firm generalizations. The object of the paper has been to encourage both forecasters and forecast users to pay more attention to estimates of forecast uncertainty.

References

Armstrong, J. Scott. 1985. *Long-Range Forecasting from Crystal Ball to Computer*. New York: John Wiley & Sons.

Braun, Phillip A., and Ilan Yaniv. 1992. "A Case Study of Expert Judgement: Economists' Probabilities versus Base-Rate Model Forecasts." *Journal of Behavioral Decision Making*, vol. 5 (March), pp. 217–31.

Chatfield, Chris. 1993. "Calculating Interval Forecasts." *Journal of Business and Economic Statistics*, vol. 11 (April), pp. 121–44.

Croushore, Dean. 1993. "Introducing: The Survey of Professional Forecasters." Federal Reserve Bank of Philadelphia *Business Review*, November/December, pp. 3–15.

Eggert, Robert J., Editor. October Issues, 1977–93. *Blue Chip Economic Indicators*. New York: Management Resources, Inc.

Engle, Robert F. 1983. "Estimates of the Variance of U.S. Inflation Based upon the ARCH Model." *Journal of Money, Credit and Banking*, vol. 15 (August), pp. 286–301.

Hafer, R.W. and Scott E. Hein. 1985. "On the Accuracy of Time Series, Interest Rate and Survey Forecasts of Inflation." *Journal of Business*, vol. 58 (October), pp. 377–98.

Hogarth, Robin M. 1980. *Judgement and Choice: The Psychology of Decision*. New York: John Wiley & Sons.

Kendall, Maurice G. 1962. *Rank Correlation Methods*, 3rd ed. New York: Hafner Publishing Co.

Lahiri, Kajal and Christie Teigland. 1987. "On the Normality of Probability Distributions of Inflation and GNP Forecasts." *International Journal of Forecasting*, vol. 3, pp. 269–79.

Lahiri, Kajal, Christie Teigland and Mark Zaporowski. 1988. "Interest Rates and the Subjective Probability Distribution of Inflation Forecasts." *Journal of Money, Credit and Banking*, vol. 20 (May), pp. 233–48.

Lahiri, Kajal and Mark Zaporowski. 1987. "More Flexible Use of Survey Data on Expectations in Macroeconomic Models." *Journal of Business and Economic Statistics*, vol. 5 (January), pp. 69–76.

McNees, Stephen K. 1992. "How Large Are Economic Forecast Errors?" *New England Economic Review*, July/August, pp. 25–42.

Tversky, Amos, and Daniel Kahneman. 1974. "Judgement Under Uncertainty: Heuristics and Biases." *Science*, vol. 185 (September), pp. 1124–31.

Zarnowitz, Victor. 1969. "The New ASA–NBER Survey of Forecasts by Economic Statisticians." *American Statistician*, vol. 23, pp. 12–16.

———. 1985. "Rational Expectations and Economic Forecasts." *Journal of Business and Economic Statistics*, vol. 3 (October), pp. 293–311.

———. 1992. *Business Cycles: Theory, History, Indicators, and Forecasting*. Chicago: University of Chicago Press.

Zarnowitz, Victor and Phillip Braun. 1992. "Twenty-Two Years of the NBER-ASA Quarterly Economic Outlook Surveys: Aspects and Comparisons of Forecasting Performance." *National Bureau of Economic Research Working Paper* no. 3965, January.

Zarnowitz, Victor and Louis A. Lambros. 1987. "Consensus and Uncertainty in Economic Prediction." *Journal of Political Economy*, vol. 95 (June), pp. 591–621.