Transparency, Expectations, and Forecasts

ANDREW BAUER, ROBERT A. EISENBEIS, DANIEL F. WAGGONER, AND TAO ZHA

Bauer is a senior economic analyst in the macropolicy section, Eisenbeis is executive vice president and director of research, Waggoner is a research economist and assistant policy adviser in the financial section, and Zha is a research economist and policy adviser in the macropolicy section, all in the Atlanta Fed's research department. *They thank Jinill Kim, Brian Madigan, John Robertson, and Ellis Tallman for critical comments and Cindy Soo and Eric Wang for research assistance. A similar version of this research is also published with the same title as Federal Reserve Bank of Atlanta Working Paper 2006-3.*

Many macroeconomists have argued that a central bank should be transparent about its objectives, its views about the economic outlook, and the reasoning behind its policy changes (see Faust and Leeper 2005). In 1994 the Federal Open Market Committee (FOMC) began to release statements accompanying changes in the federal funds rate target. Since then, the degree of specificity of the statements and the guidance provided on the likely course of future policy have evolved significantly.[1] In a recent paper, Woodford (2005) discusses two kinds of central-bank communications: current policy decisions and the central bank's view of likely future policy. He articulates four categories of information—the central bank's view of current economic conditions, current operating targets, strategies guiding policy decision making, and the outlook for future policy—that a central bank might seek to communicate to the public. Woodford argues that these open communications are "beneficial, not only from the point of view of reducing the uncertainty with which traders and other economic decision makers must contend, but also from that of enhancing the accuracy with which the FOMC is able to achieve the effects on the economy that it desires, by keeping the expectations of market participants more closely synchronized with its own."

This article investigates whether the public's views about the economy's current path and about future policy have been affected by changes in the Federal Reserve's communications policy as reflected in private-sector forecasts of future economic conditions and policy moves. In particular, has private agents' ability to predict the direction of the economy improved since 1994, when the FOMC began to publicly state its views of the economic outlook? If so, on which dimensions has the ability to forecast improved? The analysis focuses on both the short-term and longer-term economic forecasts of key macroeconomic variables—such as inflation, gross domestic product (GDP) growth, and unemployment—and of policy variables such as short-term interest rates. Private agents' current-year and next-year forecasts are used as proxies for the public's short-term and longer-term expectations, and empirical evidence is presented regarding whether such forecasts have performed better in predicting future economic and policy conditions since 1994.

The private-agent forecasts used in this article are those of individual participants as well as the consensus (average) forecasts contained in the monthly *Blue Chip Economic Indicators* surveys from 1986 to 2004, which include both the pre-FOMC-statement subperiod (1986:01–1993:12) and the post-FOMC-statement subperiod (1994:01–2004:12). We employ the econometric methodology of Eisenbeis, Waggoner, and Zha (2002), which permits us to evaluate the accuracy of forecasts both in cross section and across time and to examine the errors in forecasting key economic variables on both a univariate and a multivariate basis. The latter is important because agents are not simply forecasting one economic variable but rather a set of variables that presumably are interrelated and jointly capture important dimensions of economic performance. Good forecasts on one dimension but poor overall performance may provide some indication of the internal consistency of the forecaster's approach.

This cross-sectional data set enables us to decompose forecast accuracy into two components: the common error that affects all individual participants and the idiosyncratic error that reflects discrepant views across individuals about future economic and policy conditions. According to Woodford (2005), one should expect the idiosyncratic error to become smaller as FOMC open communications become more transparent. But the common error may not change much because it is likely to be affected by factors other than changes in policy transparency, such as unforeseen business cycles.

To preview the main result, we find that since 1994 the idiosyncratic errors for key macroeconomic variables have steadily declined and the expectations of market participants are more closely synchronized to one another. We find no evidence, however, that the common error has become smaller since 1994, especially for the longer-term forecasts.

The Methodology

Let μ_t be an $n \times 1$ vector of economic variables at time t, let y_t be the realized value of these economic variables, and let y_t^i be the ith individual's forecast value of the variables. Assume that y_t is normally distributed with mean μ_t and an economywise (common) covariance matrix Ω_t^R and that y_t^i is normally distributed with mean μ_t and a forecastwise covariance matrix Ω_t^F. (The superscripts R and F stand for "realized" and "forecast," respectively.) The covariance matrix Ω_t^R reflects the aggregate shocks that affect the realized value of μ_t; the covariance matrix Ω_t^F captures the discrepancy in forecasts across individual participants. The assumption that the mean forecast among individual participants is μ_t is reasonable because previous work has suggested that the Blue Chip Consensus forecast, serving as a proxy for the mean forecast, is close to being an unbiased estimate of μ_t (Bauer et al. 2003). We denote the forecast error for the ith forecaster by $x_t^i = y_t^i - y^t$. Therefore, the individual forecast error x_t^i has mean zero and a variance matrix

$$\Omega_t = \Omega_t^R + \Omega_t^F,$$

which indicates that x_t^i is subject to both idiosyncratic and common shocks.[2] The standard statistical theory implies that

$$\chi_t^i \equiv x_t^{i\prime} \Omega_t^{-1} x_t^i \sim \chi^2(n),$$

where $\chi^2(n)$ denotes the χ^2 distribution with n degrees of freedom and χ_t^i is a square error weighted by Ω_t. The above expression shows that the weighted square error

χ_t^i follows the χ^2 distribution with n degrees of freedom. To measure the forecast accuracy for each individual participant, we compute a score value (p value) associated with this χ^2 distribution and call it an "accuracy score." The score for individual forecaster i at forecast time t is a function of χ_t^i and n:

$$p\left(\chi_t^i, n\right) = 1 - \chi_{cdf}^2\left(\chi_t^i, n\right),$$

where $\chi_{cdf}^2(\chi_t^i, n)$ is the probability that a random observation from the χ^2 distribution with n degrees of freedom falls in the interval $[0\ \chi_t^i]$.[3]

As Eisenbeis, Waggoner, and Zha (2002) point out, the summary measure $p(\chi_t^i, n)$ is a probability that is invariant to the underlying scales-of-error variances. One possible interpretation is that the ith participant's forecast is closer to the realized value than that of 100 $p(\chi_t^i, n)$ percent of all possible forecasters. Moreover, the score $p(\chi_t^i, n)$ can be compared across forecasters, within a forecast period, and across periods.

Bauer et al. (2003) show how to estimate the covariance matrices Ω_t^R and Ω_t^F. The matrix Ω_t^R can be estimated as the sample covariance matrix of the Blue Chip Consensus forecast errors across time under the assumption that Ω_t^R is the same across years for each month but varies across months within a year. Thus, the variances on the diagonal of Ω_t^R become smaller as t approaches the end of the year because more information becomes available to forecast economic conditions for the current year. The covariance matrix Ω_t^F can be estimated as the sample covariance matrix of forecast errors across individual forecasters; this covariance varies both across months and across years.[4] The estimate of Ω_t, denoted by $\hat{\Omega}_t$, is the sum of the estimates of Ω_t^R and Ω_t^F. Given this estimate, the weighted-square error can be calculated as

$$\hat{\chi}_t^i = x_t^{i\prime} \hat{\Omega}_t^{-1} x_t^i.$$

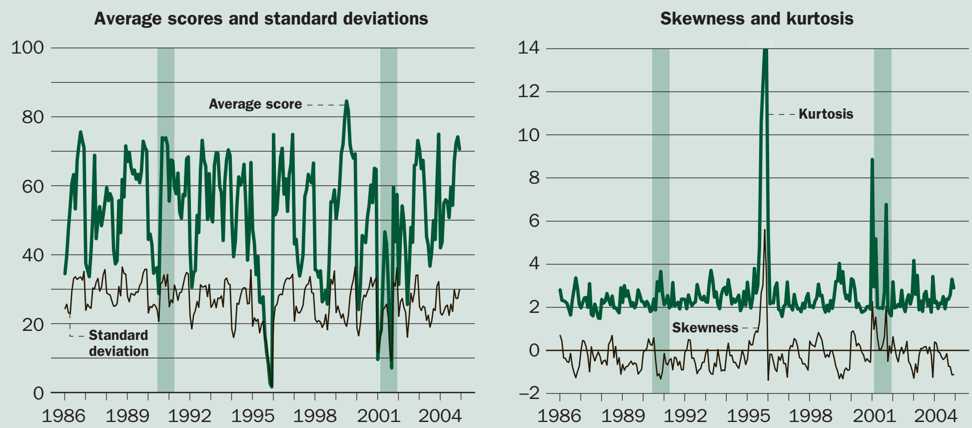
At each time t, the average accuracy score is

$$\hat{p}_t(n) = \frac{1}{N_t} \sum_{i=1}^{N_t} p\left(\hat{\chi}_t^i, n\right),$$

where N_t is the number of individual forecasters at time t. One can also calculate the cross-sectional distribution of accuracy scores; the process is described in detail in the sidebar on page 6.

1. Kohn and Sack (2003) characterize several distinct periods of increasing transparency in FOMC statements: statements on changes in the discount rate (1989–93), statements on changes in the federal funds rate (1994–98), statements including policy tilt (1998–99), and statements including assessment of the balance of risks (2000–04). In May 2003 a further refinement was added to separately state the committee's views on the risks to inflation and growth. And, finally, in August 2003 the committee provided explicit guidance on the likelihood that policy would remain accommodative.
2. In future research, we intend to relax the assumptions that the Blue Chip Consensus forecast is equal to μ_t and idiosyncratic shocks are independent of common shocks.
3. If the assumptions used are valid, the distribution of accuracy scores from 1986 to 2004 should be uniform. We have verified that such a distribution is more or less uniform, taking into account small-sample uncertainty.
4. Other estimates can also be constructed using model-based methods.

Figure 1
Blue Chip Average of Individual Scores for the Current Year



Note: The shaded vertical bars indicate recessions.

Source: Authors' calculations from monthly *Blue Chip Economic Indicators* data

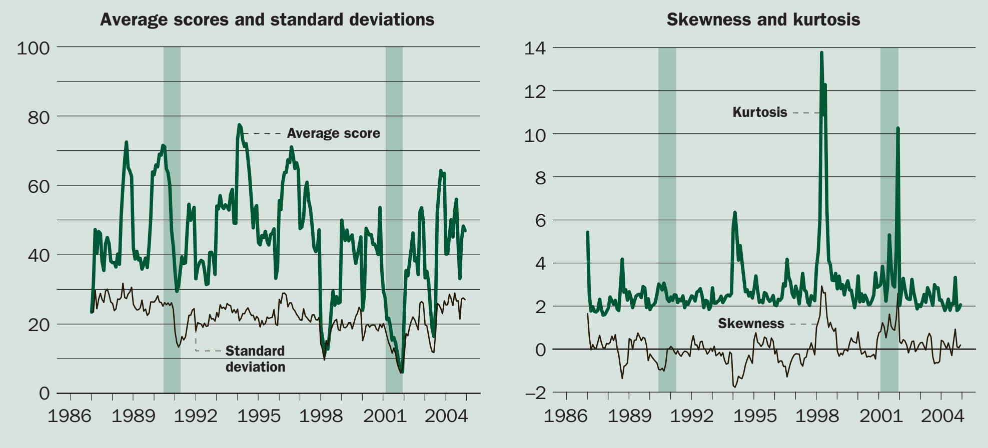
Vintage Data and Forecast Errors

The monthly *Blue Chip Economic Indicators* report the forecasts of key macroeconomic variables for the current and next years. We study the annual average forecasts of five key variables: the three-month Treasury bill (T-bill) rate, the consumer price index (CPI) inflation rate, real gross national product (GNP) for 1986 to 1995 or real gross domestic product (GDP) from 1996 to 2004, the unemployment rate, and the long-term bond yield (the corporate bond yield from 1986 to 1995 or the ten-year Treasury note yield from 1996 to 2004). The three-month T-bill rate, the CPI inflation rate, the unemployment rate, and the long-term bond yield are monthly variables while real GNP/GDP is a quarterly variable. This frequency difference is important to note when evaluating forecasts. (See Appendix 1 for a description of and sources for these data.)

More information becomes available about the actual current-year data as the end of the year approaches, and therefore the forecast errors for both the current and next years get smaller. For example, the forecasters participating in the December Blue Chip survey will have monthly data on the three-month T-bill rate and the long-term bond yield through November, data on the unemployment rate through October or November, and data on the CPI inflation rate through October. However, since GNP/GDP data are released quarterly, forecasters will have information regarding GNP/GDP only through the third quarter of the year. The weighted-square error $\hat{\chi}_t^i$ is designed to avoid the influence of different amounts of available data so that the errors are comparable across time.

To gauge forecast errors, the realized values of each variable at a given time must be used. The values of some variables are revised over time by the agencies responsible for reporting those variables. In particular, real GNP/GDP is reported quarterly and revised twice. Every year additional benchmark revisions may be made in July to past GDP data. Hence, the information reported is actually the continuously changing estimates of many key economic variables' final values. Finally, sometimes the definition of GDP is changed and the series is completely revised.

Figure 2
Blue Chip Average of Individual Scores for the Next Year



Note: The shaded vertical bars indicate recessions.

Source: Authors' calculations from monthly *Blue Chip Economic Indicators* data

Such revisions raise the question, What vintage data should one use to evaluate forecast errors? From a macropolicy perspective, one could argue that the focus should be on the "best" estimate of the final value of the variable of interest. Often, however, that value is not known for several years, and sometimes the difference between even a preliminary estimate and its nearest neighbor estimates can be very large. For example, the advanced estimate for real GDP for the first quarter of 2005 was 3.1 percent. This number was revised upward by the Bureau of Economic Analysis (BEA) to 3.4 percent and finally to 3.8 percent as more data on the performance of the economy became available. Policymakers might have inferred that the economy was growing below trend according to the first number but above trend based on the final estimate. Such differences could have significantly different implications for policy. For this reason, we would argue that the focus should be on forecast methods that best approximate the final number rather than the initial estimate. Also, a priori knowledge of the expected performance of a model or forecasting method can help policymakers decide how to weigh the evidence when significant differences exist between the initial releases of data and forecasts.

For the purposes of this study, for the current-year forecasts, we use vintage data available at the end of January following the current year; for the next-year forecasts, we use data available at the end of January following the next year. This study uses vintage data so that its results will be comparable with those of previous studies. It also provides a comparison between the average Blue Chip Consensus score using vintage and final data, using January 2005 for the final data.

Accuracy Scores

This section looks at the distribution of scores at each month and examines whether the distribution has changed over time, especially from the prestatement subperiod to the poststatement period. The technical details of how to characterize the cross-sectional distribution of scores are provided in the sidebar on page 6.

The first panel of Figure 1 shows the time-series paths of average scores and standard deviations of scores for the current year. The first panel of Figure 2 shows

Characterizing the Distribution of Accuracy Scores

The distribution of accuracy scores can be summarized by the first four moments. The method for calculating the mean or average score $\hat{p}_t(n)$ is shown in the text. The other three moments—standard deviation, skewnesss, and kurtosis—can be calculated as follows:

$$\hat{\sigma}_t(n)=\left[\frac{1}{N_t}\sum_{i=1}^{N_t}\left(p\left(\hat{\chi}_t^i,n\right)-\hat{p}_t(n)\right)^2\right]^{\frac{1}{2}},$$

$$\hat{s}_t(n)=\frac{\frac{1}{N_t}\sum_{i=1}^{N_t}\left(p\left(\hat{\chi}_t^i,n\right)-\hat{p}_t(n)\right)^3}{\hat{\sigma}_t(n)^3}, \text{ and}$$

$$\hat{u}_t(n)=\frac{\frac{1}{N_t}\sum_{i=1}^{N_t}\left(p\left(\hat{\chi}_t^i,n\right)-\hat{p}_t(n)\right)^4}{\hat{\sigma}_t(n)^4},$$

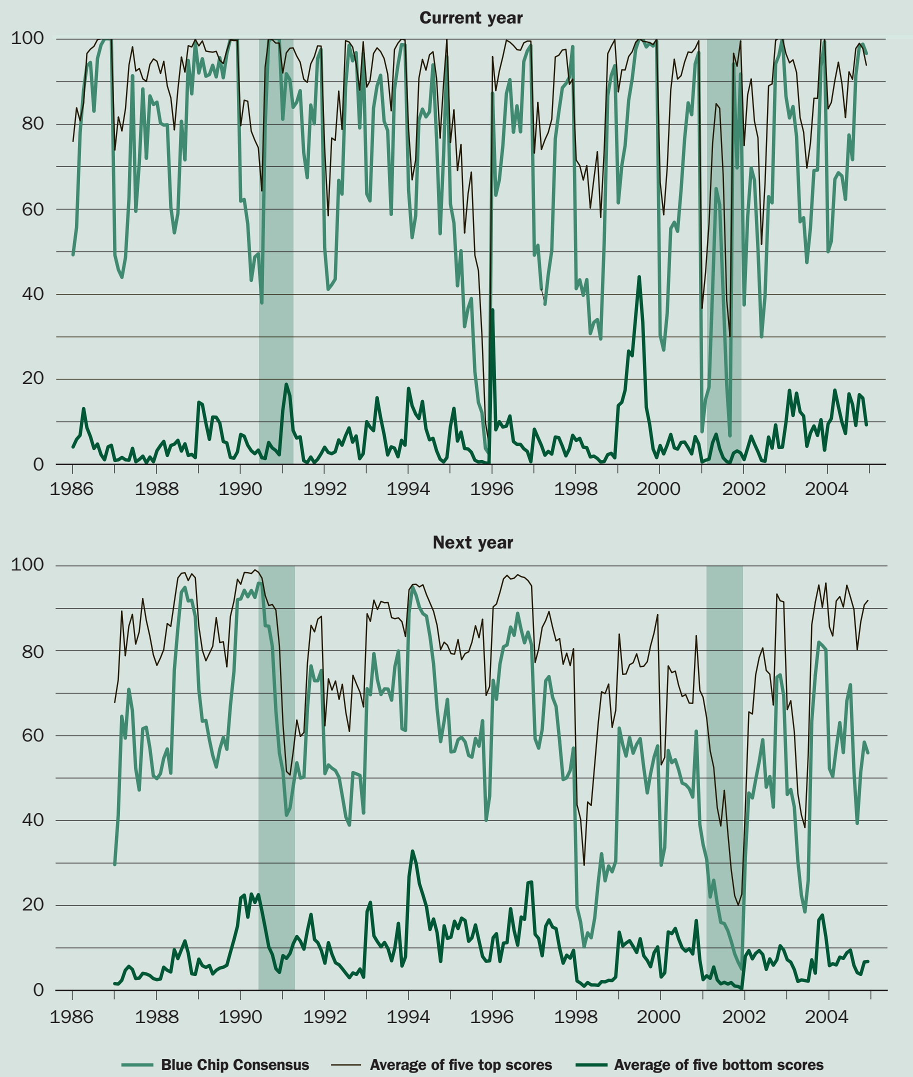
where σ stands for the standard deviation, s the skewness, and u the kurtosis.

similar paths for the next year. The measure of standard deviation is often used to approximate the volatility of the public's expectations or forecasts at each point in time. As the first panel of Figure 1 shows, both the average score and the standard deviation of scores fluctuate over time. No noticeable differences exist in the degree of fluctuation before and after 1994, nor are there differences for any subperiods after 1994. No trend appears in which the average score has increased or the standard deviation of scores has decreased since 1994. The figures clearly display periods when forecasters made big errors, such as missing the onset of the recessions in 1990 and 2001. In addition, while the average scores increased in 2004, so did the standard deviations of the scores. Similarly, the average scores dropped significantly in 1995 primarily because the definition of the GDP series changed. In January 1996 the BEA changed the measurement of GDP to a chain-weighted system, but the forecasts made before January 1996 might be based on the non-chain-weighted series. Interestingly, this change seems to have had relatively less effect on the longer-term forecast errors (the second panel of Figure 2).

The average score for the next year (Figure 2) shows no improvement since 1994 and in fact appears to have drifted lower since 1996. The standard deviation of scores since 2001 has drifted steadily upward. The pattern of the drift in the standard deviation is similar to that just prior to and coming out of the 1990–91 recession. As discussed further in the next section, these lower scores after 1996 are most likely associated with the nature of the business cycle and a surge of unexpected productivity growth in the late 1990s.

The second panels of Figures 1 and 2 display the skewness and kurtosis of accuracy scores. Skewness measures the asymmetry of the score distribution. The more negative this measure is, the more scores spread out toward 0 percent. Conversely, the more positive this measure is, the more scores spread out toward 100 percent. Kurtosis measures the likelihood that the score distribution has extreme outliers that may affect the average score. The bigger the value of this measure is, the more likely the presence of outliers in the score distribution is. For the current-year forecasts, the skewness and kurtosis have remained stable except for a few periods. The 1995 spike is the result of the redefinition of GDP, and the small spikes around 2001 are associated with the recent recession. For the next-year forecasts, again, no clear pattern or trend is apparent in which skewness and kurtosis have changed since 1994. Two spikes in skewness and kurtosis correspond to the Asian financial crisis and the recent recession.

Figure 3
Blue Chip Consensus Scores and the Averages of the Five Top and Bottom Forecaster Scores

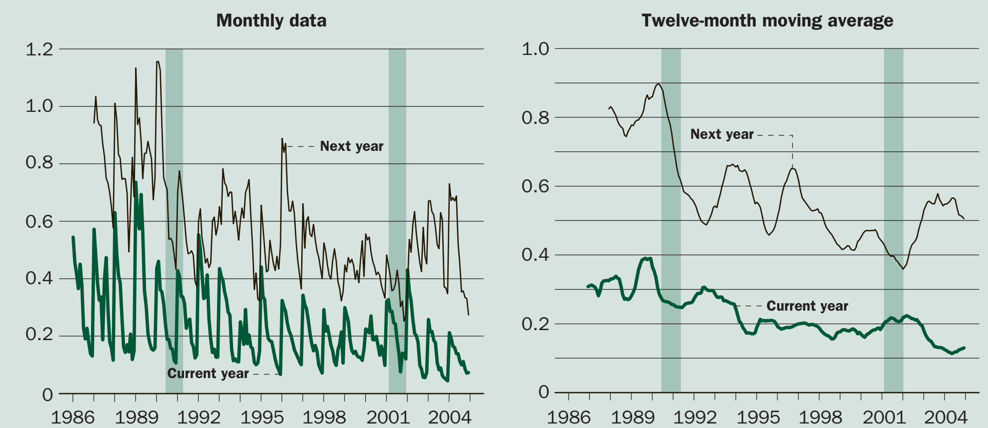


Note: The shaded vertical bars indicate recessions.

Source: Authors' calculations from monthly *Blue Chip Economic Indicators* data

Further information about the distributional changes of accuracy scores is provided in Figure 3, which displays the time-series paths of accuracy scores of the Blue Chip Consensus forecast and the average of the top and bottom five forecasts for each month. The consensus forecast is of particular interest because its score is on average the highest (see Appendix 2 for details) and because it performs better than any single individual forecaster over the sample. Again, Figure 3 demonstrates that these scores have had no tendency to improve over time since 1994. In fact, the

Figure 4
Cross-Sectional Standard Deviations of Three-Month Treasury Bill Forecasts



Note: The shaded vertical bars indicate recessions.

Source: Authors' calculations from monthly *Blue Chip Economic Indicators* data

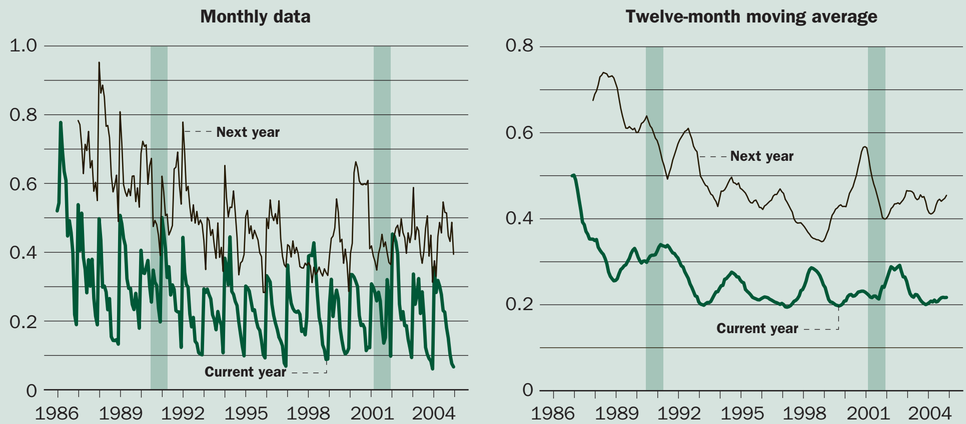
scores of consensus forecasts appear to be slightly lower after 1996 than before, especially for the next-year forecast. Moreover, the drop in the consensus scores around the recent recession and again following September 11, 2001, suggests that events and exogenous shocks affected forecast performance much more than FOMC statements did. The drop in the scores toward the end of 1995 is attributable to the redefinition of GDP. The average scores for the five top and the five poorest forecasters suggest that the data have fat tails, with most of the forecasts being clustered at the high end with a few really poor performers on the bottom.

All these findings suggest that the individual participant's forecast performance relative to other participants has not improved between the prestatement and poststatement periods. Although the accuracy score is a powerful summary measure of forecasting performance, it is a nonlinear function of the square forecast errors weighted by the overall covariance matrix Ω_t. Separating Ω_t and forecast errors for further analysis would be informative. In the next section, we examine whether the covariance matrix Ω_t^F has changed over time and study the sources of forecast errors that do not depend on Ω_t^F and Ω_t^R.[5]

Transparency and Sources of Forecast Errors

Kohn and Sack (2003) and Woodford (2005) argue that the contents of FOMC statements have become more transparent since 1994. To evaluate this argument, it is important to determine whether the expectations of market participants as reflected in the forecasts of key economic variables have become more synchronized in the poststatement period than in the prestatement subperiod. If the statement contains useful information, then one might expect an overall improvement in forecast accuracy, ceteris paribus, or at least more agreement among forecasters (that is, a tighter distribution of idiosyncratic errors). A positive answer may provide evidence about the effects of the FOMC statements on the private sector's agreement on the direction of the future economy.

Figure 5
Cross-Sectional Standard Deviations of CPI Forecasts



Note: The shaded vertical bars indicate recessions.

Source: Authors' calculations from monthly *Blue Chip Economic Indicators* data

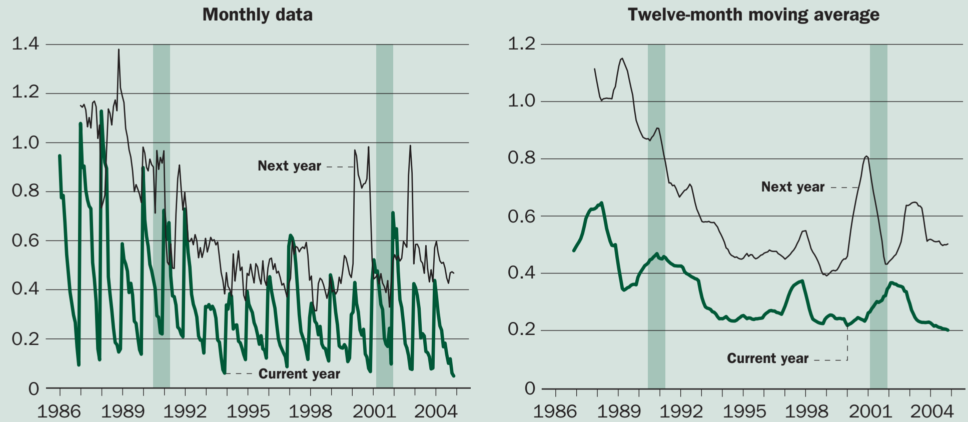
We also examine the sources of forecast errors by directly decomposing the mean square error (MSE) into the idiosyncratic component that reflects the discrepancy in individual participants from the Blue Chip surveys and the common component that is associated with unanticipated aggregate shocks and affects all participants. The technical details of this decomposition are provided in the sidebar on page 19.

The MSE is the average of square errors across individual forecasters. Arguably, both the idiosyncratic and common errors may show a decreasing trend if the statement contains useful information and forecasters gain better understanding of the economy over time, especially after 1994. To the extent that the common error is affected by exogenous aggregate shocks and the distribution of the shocks is not constant, no clear inference may exist about the size of the common error. However, we hypothesize that the more important impact is likely to be seen for the idiosyncratic component, in that the idiosyncratic errors should be tighter—that is, greater agreement should be evident among the forecasters. The empirical results presented below confirm this hypothesis.

The degree of synchronization among market participants' expectations is measured by the cross-sectional standard deviations of all the variables, which are equal to square roots of the diagonal elements of Ω_t^F. Figures 4–8 report the cross-sectional standard deviation of each of the five macroeconomic variables considered in this study. These charts clearly show that the trend for these variables has been downward, and the standard deviations tend to be smaller after 1994 than before 1994. These findings suggest that individual participants' forecasts have indeed been more synchronized since 1994 in terms of both their overall view of the economy and the interest rate variable most closely tied to policy.

5. The reader may recall that by assumption Ω_t^R does not change from one year to another. We intend to relax this assumption in future research.

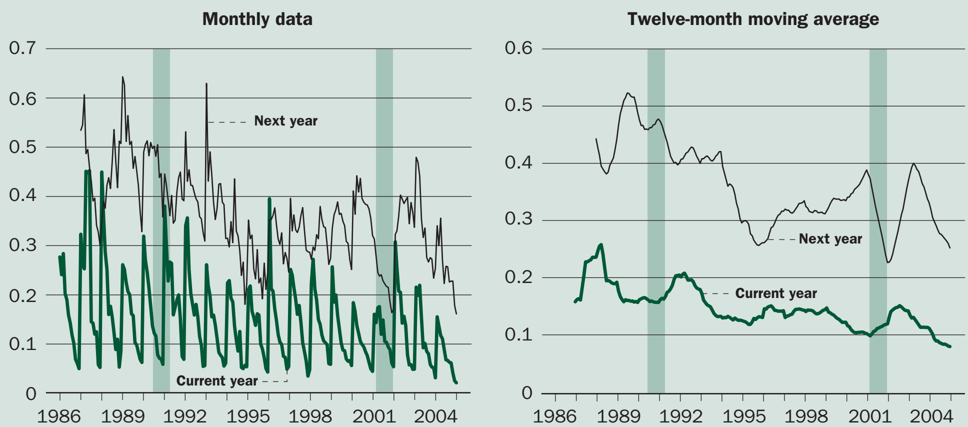
Figure 6
Cross-Sectional Standard Deviations of GDP Forecasts



Note: The shaded vertical bars indicate recessions.

Source: Authors' calculations from monthly *Blue Chip Economic Indicators* data

Figure 7
Cross-Sectional Standard Deviations of Unemployment Rate Forecasts

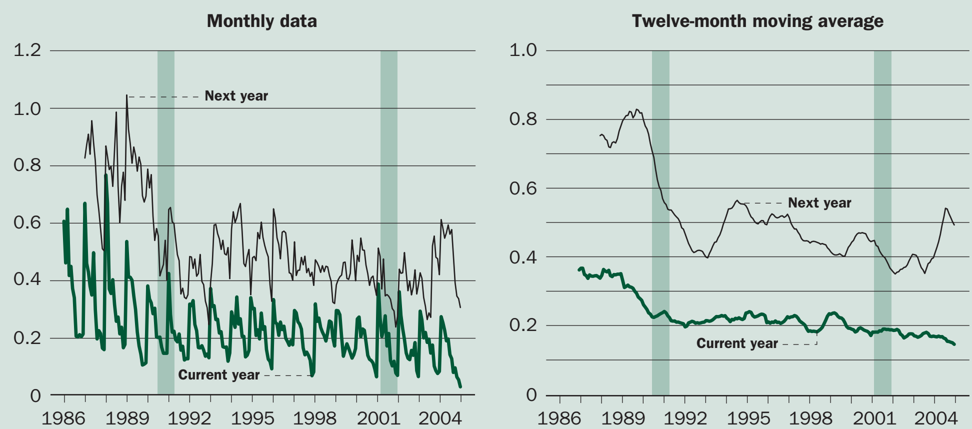


Note: The shaded vertical bars indicate recessions.

Source: Authors' calculations from monthly *Blue Chip Economic Indicators* data

Figures 9–14 show the time-series paths of decompositions for each of the five key variables as well as all the variables jointly. One uniform result seen in the first panel of each figure is that the time path of idiosyncratic errors shows a pattern of steady decline as well as a seasonal pattern for the current-year forecasts. Within the current year, the individual participant's forecast error becomes much smaller as December approaches. The seasonal pattern is much less obvious for the next-year forecasts (the second panel of each figure) partly because the uncertainty about the economy during the coming year is still large even if one tries to forecast as of

Figure 8
Cross-Sectional Standard Deviations of Ten-Year Treasury Note Forecasts



Note: The shaded vertical bars indicate recessions.

Source: Authors' calculations from monthly *Blue Chip Economic Indicators* data

December in the current year. For both the current-year and next-year forecasts, a clear pattern of smaller idiosyncratic errors emerges after 1994. Again, these results are consistent with the hypothesis that individual forecasts have been more synchronized since 1994.

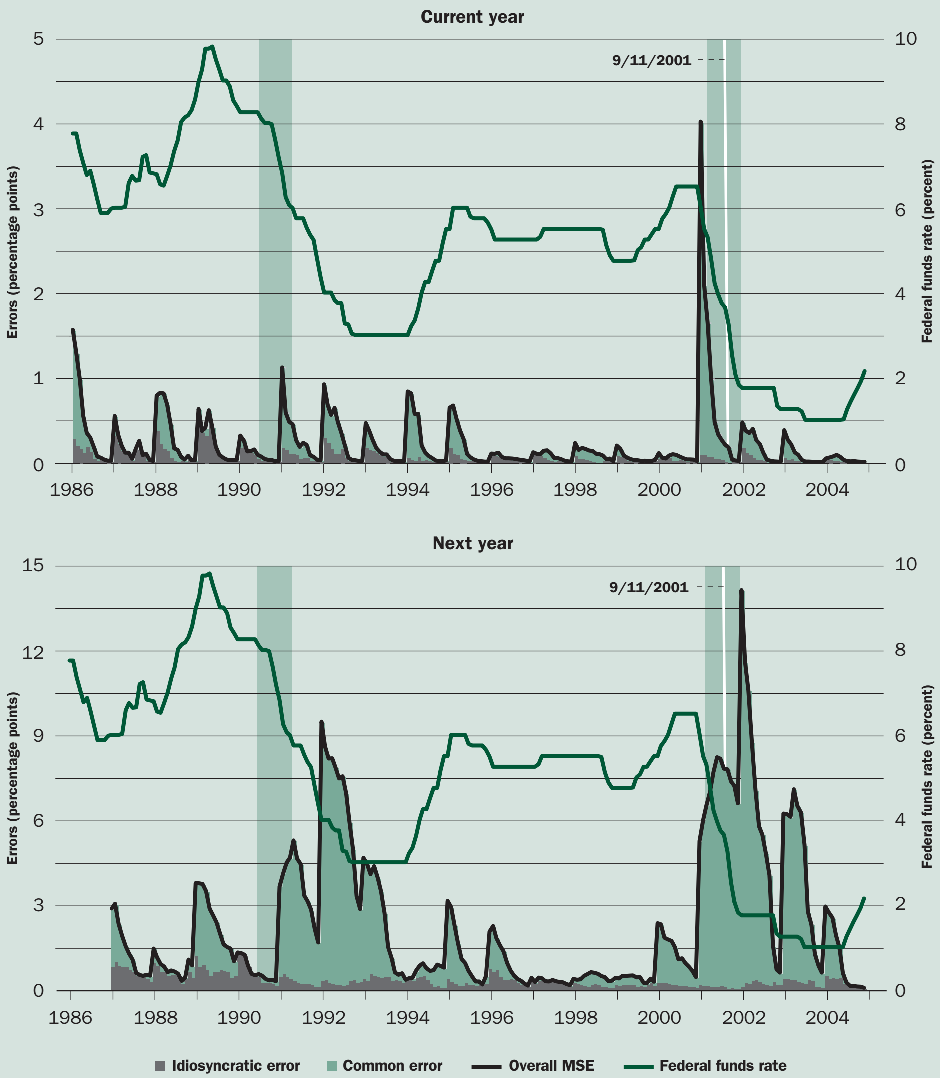
Patterns of common errors are distinctively different from those of idiosyncratic ones, and the difference seems to be associated with business cycles unrelated to the FOMC statements. One can see from Figures 9–14 that the common errors in the current-year forecast are large relative to the idiosyncratic errors whereas the common errors are dominant in the next-year forecasts. But there is no apparent pattern that the common errors are smaller after 1994 than before.

According to the first panel of Figure 9, unusually large common errors for the current-year forecasts of the short-term interest rate occur in 2001. These errors are associated with the unexpected sharp decline of the federal funds rate. The large common errors of longer-term (next-year) forecasts seem to be associated with missing the turning point of the federal funds rate in the early 2000s and failing to predict the unchanged rate in 2002 and 2003 (the second panel of Figure 9).

For CPI inflation, except for two unusually large common errors before 1994, the common errors of the current-year forecasts have similar patterns before and after 1994 (the first panel of Figure 10). The common errors for the next-year forecasts tend to be larger in the period after 1996 than before (the second panel of Figure 10), and no tendency is apparent that these errors have become smaller than before 1994.

Typically, as the end of the year approaches, both idiosyncratic and common errors become smaller for the current-year forecasts. But unusually large common errors of the current-year forecasts of real GNP/GDP develop toward the end of 1995, caused mainly by the definition change of the GDP series. When divided by the diminishing variances of forecast errors, these errors are amplified, accounting for the steep drop of accuracy scores toward the end of 1995 (see the first panel of Figure 3). In the first panel of Figure 11, the errors are not divided by the variances of forecast errors and thus are not as visually dramatic as in Figure 3. The substantial,

Figure 9
Mean Square Errors of Three-Month Treasury Bill Forecasts



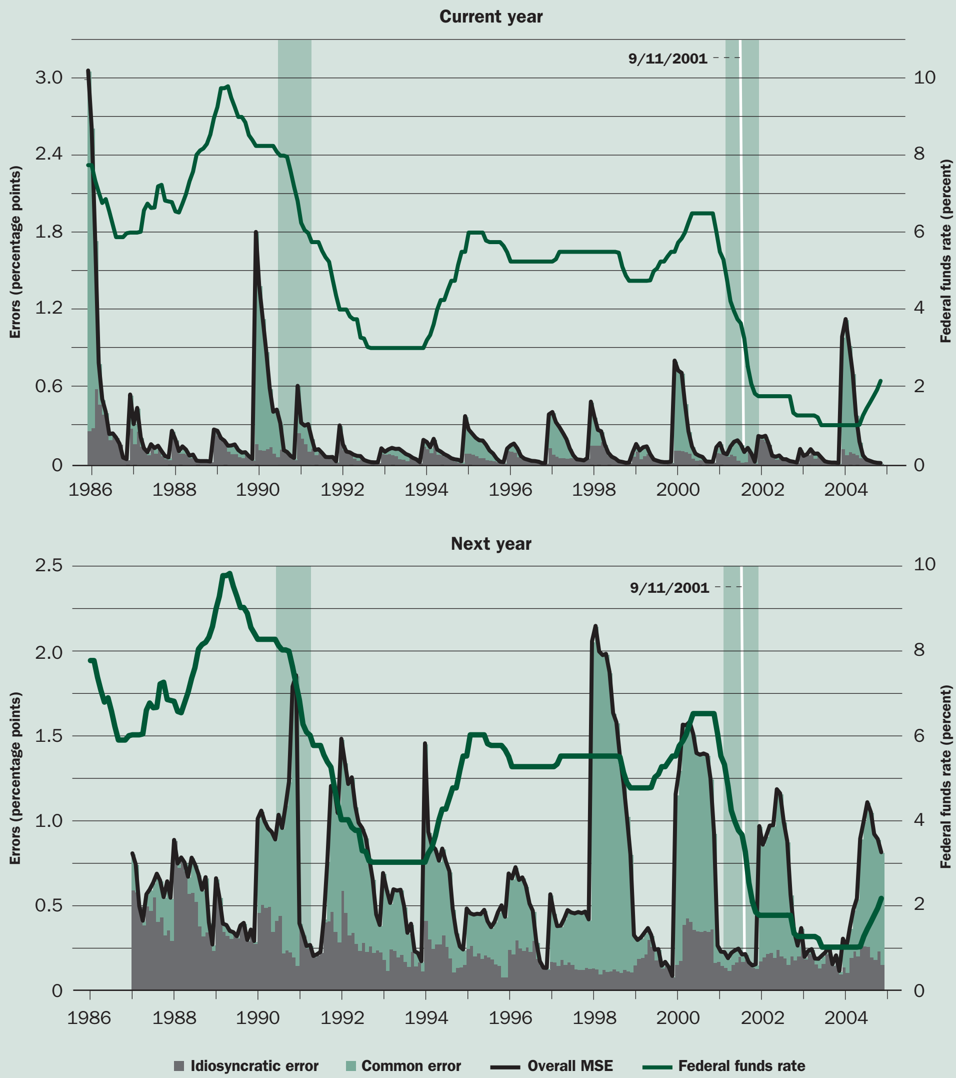
Note: The shaded vertical bars indicate recessions.

Source: Authors' calculations from monthly *Blue Chip Economic Indicators* data

persistent common errors of the next-year forecasts in the late 1990s are consistent with the sustained increase in productivity growth being largely unexpected by the public, while the federal funds rate did not change much.

The common errors in forecasting the unemployment rate for the current year appear to be somewhat smaller after 1994 than before, but those errors for the next year have similar patterns before and after 1994 (Figure 12). The large common errors for the next-year forecasts have much to do with business cycles and with the errors in predicting output growth.

Figure 10
Mean Square Errors of CPI Forecasts



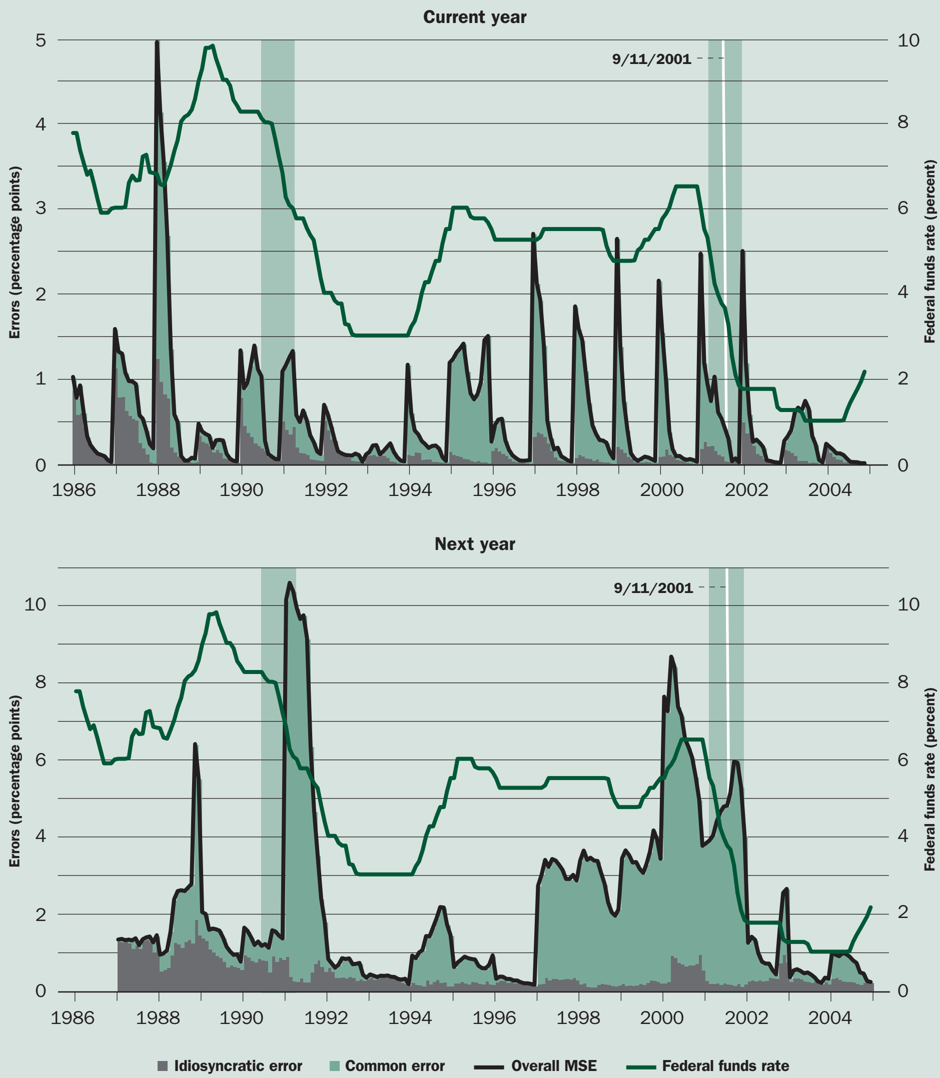
Note: The shaded vertical bars indicate recessions.

Source: Authors' calculations from monthly *Blue Chip Economic Indicators* data

No clear patterns exist in which the common forecast errors of the long-term bond yield have become smaller since 1994 (Figure 13). In particular, the errors around the recent recession are relatively large in magnitude. Interestingly, a noticeable drop in the idiosyncratic errors in both the current-year and next-year forecasts occurs after 1987, when Alan Greenspan became chairman and the effects of the stock-market problems dissipated.

Figure 14 summarizes the decomposition of the MSE for the five variables combined. For the current-year forecasts, the seasonal pattern is evident, as explained

Figure 11
Mean Square Errors of GDP Forecasts



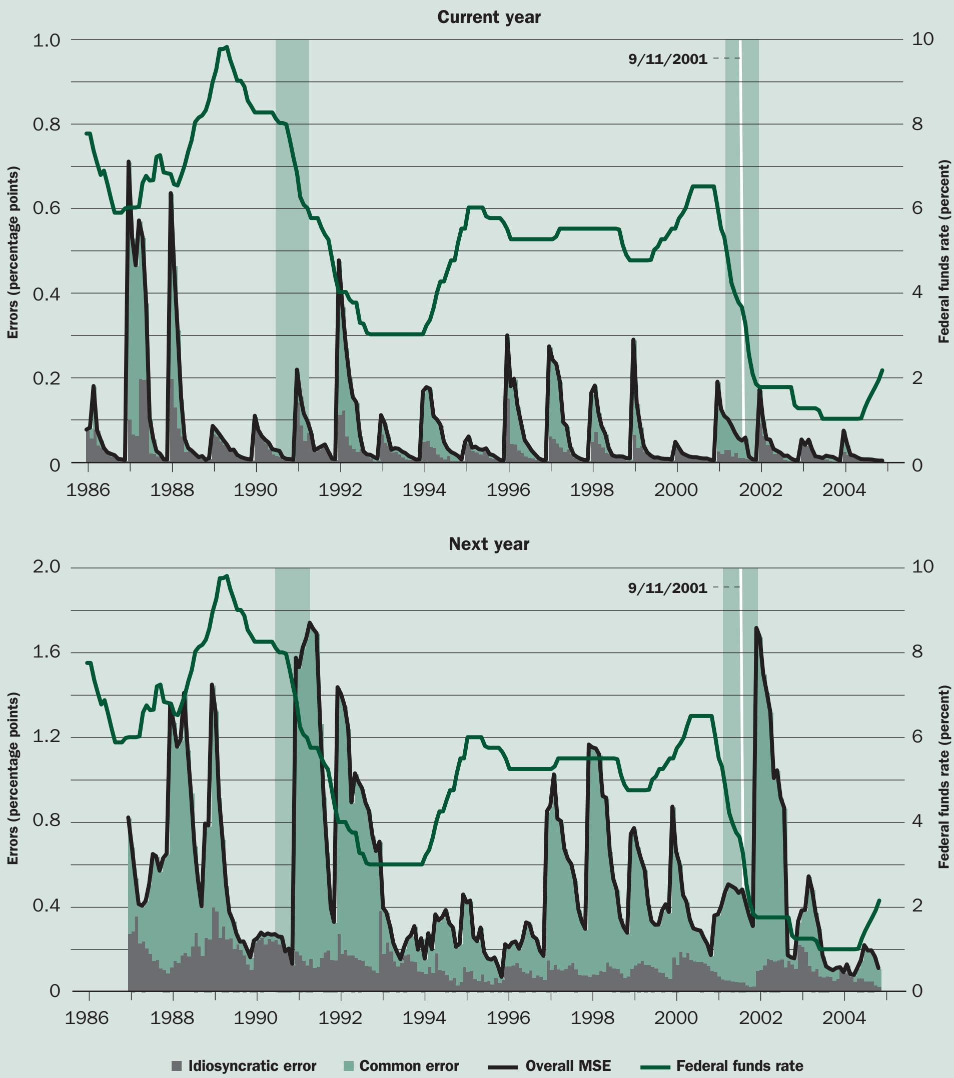
Note: The shaded vertical bars indicate recessions.

Source: Authors' calculations from monthly *Blue Chip Economic Indicators* data

early in this article. For the next-year forecasts, the large common errors occurred in the periods around the last two recessions. The persistent and volatile common errors since 1994 are mainly caused by the correlation effect among forecast errors across variables because the forecast errors for individual variables other than GNP/GDP do not share these features. Overall no evidence indicates that the public's forecasts of key macroeconomic variables have improved since 1994, following the FOMC's efforts to increase transparency.

The table (on page 18) reports the average of percentages of the MSE that are attributed to the idiosyncratic component and the common component. Two meth-

Figure 12
Mean Square Errors of Unemployment Rate Forecasts



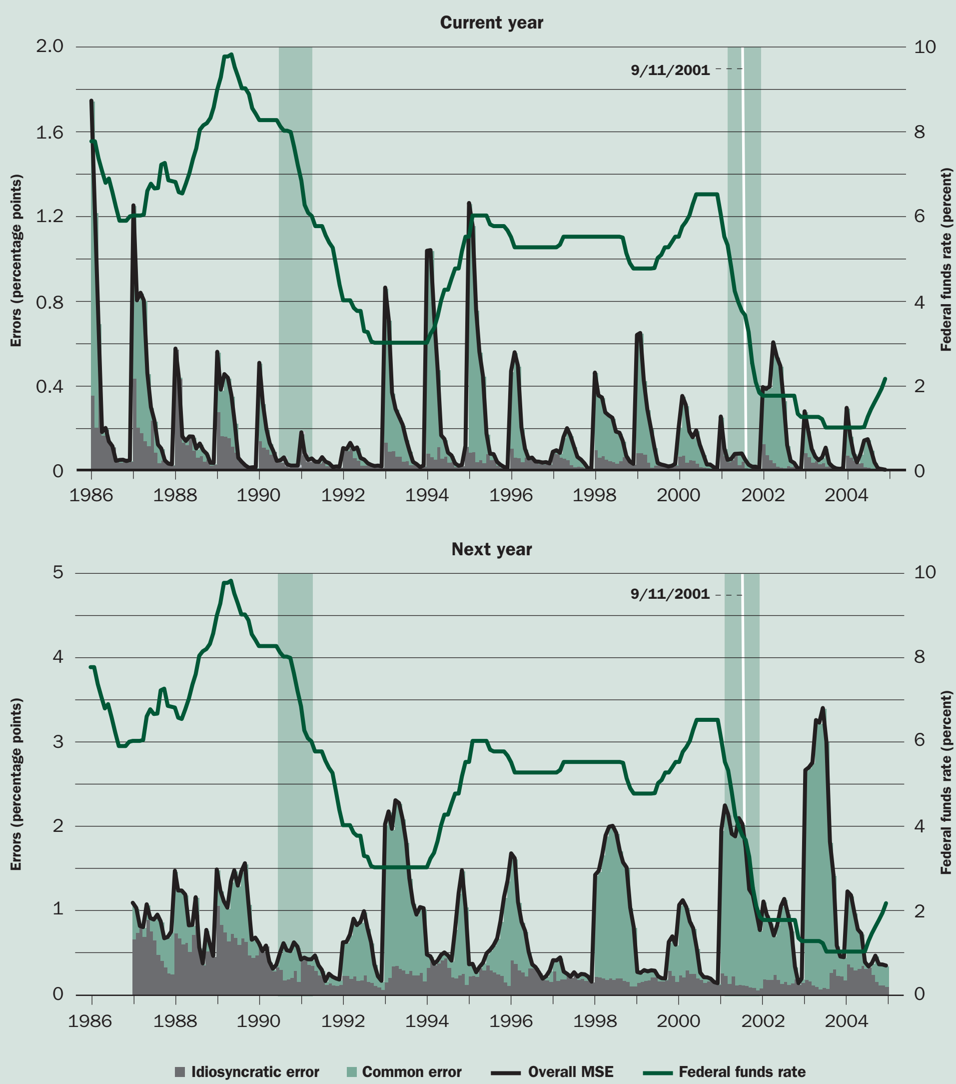
Note: The shaded vertical bars indicate recessions.

Source: Authors' calculations from monthly *Blue Chip Economic Indicators* data

ods are used to compute the average percent contributions. The first is to calculate the percent contributions of idiosyncratic and common errors for each period and then average them over all the periods. This method helps eliminate outliers of extremely large errors, so the results may not conform to the patterns in the charts. The top panel of the table reports these results.

The second method is to accumulate the forecast errors of both types throughout the entire sample and then calculate the percent contributions of idiosyncratic and common errors (see the bottom panel of the table). This method is likely to be influenced by outliers but will be consistent with the patterns shown in the charts.

Figure 13
Mean Square Errors of Ten-Year Treasury Note Forecasts

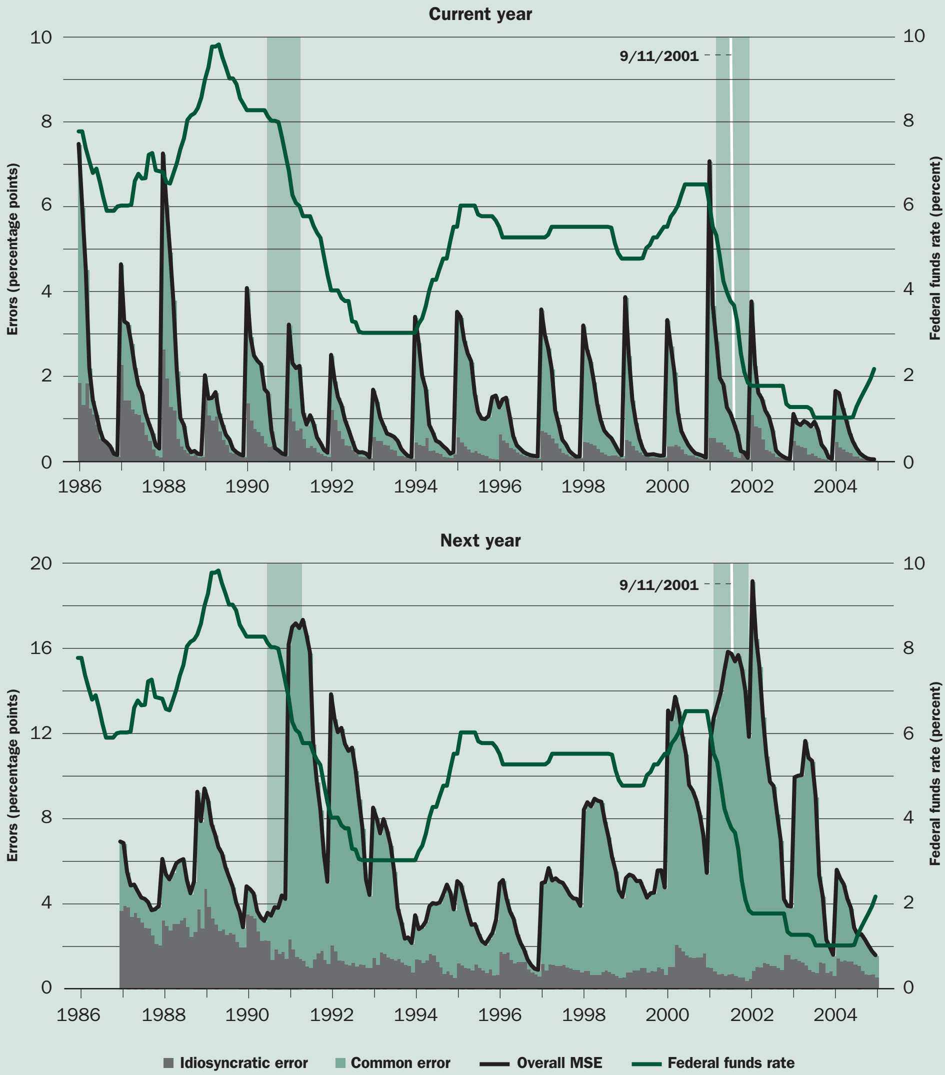


Note: The shaded vertical bars indicate recessions.

Source: Authors' calculations from monthly *Blue Chip Economic Indicators* data

In the top panel of the table, the idiosyncratic errors for the current-year forecasts, except for GNP/GDP, contribute much more to the total errors than the common errors do despite the fact that the common errors are much larger at times. But for all the variables jointly, the common errors become more important. This result implies that while predicting a single variable may be relatively easy, predicting a set of economic variables may be more difficult.[6] For the longer-term (next-year) forecasts, the picture is completely different: The common errors are clearly a driving force for almost all variables (except for CPI), individually and jointly.

Figure 14
Mean Square Errors of All Variables Forecasts



Note: The shaded vertical bars indicate recessions.

Source: Authors' calculations from monthly *Blue Chip Economic Indicators* data

Compared to the results in the top panel of the table, the results in the bottom panel give a more dominant role to the common errors, partly because the common errors are much larger than the idiosyncratic errors in some periods. All in all, the common errors clearly play a dominant role in overall forecast errors.

6. One might also infer that different models are being used and that these models perform better on some variables than others, but in aggregate significant differences exist among the forecasts.

Table
Decomposition of the Mean Square Error

	All variables	3-month T-bill	CPI	GDP	Unempl. rate	10-year T-note
By average percent contribution to error in each period						
Current-year forecasts (1986–2004)						
Idiosyncratic component	44.5	57.0	69.7	43.3	64.0	58.7
Common component	55.5	43.0	30.3	56.7	36.0	41.3
Next-year forecasts (1986–2003)						
Idiosyncratic component	30.0	40.0	52.7	41.0	36.6	48.5
Common component	70.0	60.0	47.3	59.0	63.4	51.5
By percent contribution of total error across sample						
Current-year forecasts (1986–2004)						
Idiosyncratic component	31.9	30.9	40.6	28.0	39.6	32.0
Common component	68.1	69.1	59.4	72.0	60.4	68.0
Next-year forecasts (1986–2003)						
Idiosyncratic component	22.1	15.1	38.6	20.1	24.7	32.1
Common component	77.9	84.9	61.4	79.9	75.3	67.9

This finding suggests that unexpected shocks, which of course are also not anticipated in the FOMC statements, are dominant factors in affecting forecast performance, and improvements in policy transparency would be unlikely to make the forecast errors smaller except on the margins.[7] Another possibility is that clearer patterns may show up as more observations become available; the FOMC only began in August 2003 to provide explicit guidance on the likely path of future policy and state-contingent economic conditions in the future. Given the data available today, however, we find no empirical evidence of significant improvement in the common forecast errors over the period in which the FOMC attempted to clarify its views of the economy or the likely course for future policy. This finding does not necessarily suggest that the movement toward transparency has been a failure. It may simply indicate that no new information was provided in the statements that had not already been inferred by market participants. Given the unpredictable nature of business cycles, moreover, the common error may be mostly affected by factors other than monetary policy transparency.

Vintage Data versus Final Data

One could argue that whenever forecast errors for a particular period are evaluated, final data available at that time should be used. The reason is obvious: From a policy perspective, being able to accurately predict initially released data that are subsequently revised may lead to policy errors, especially when turning points are imminent or when the revisions may substantially alter one's view of the economy. However, when policy formulation relies heavily upon model forecasts, it is important that those forecasts capture, as well as possible, the true underlying paths for key economic variables. If they do not, then the risk of serious policy errors may be increased. Furthermore, deciding how to choose the vintage data at various points in time is completely arbitrary, and no statistical or economical foundation exists to

Decomposition of the Mean Square Error

Let the estimate of μ_t be

$$\hat{\mu}_t = \frac{1}{N_t}\sum_{i=1}^{N_t} y_t^i.$$

Note that $\hat{\mu}_t$ is also the Blue Chip Consensus forecast. The weighted mean square error at time t can be decomposed as

$$\frac{1}{N_t}\sum_{i=1}^{N_t} x_t^{i\prime} x_t^i = \frac{1}{N_t}\sum_{i=1}^{N_t}\left[\left(y_t^i - \hat{\mu}_t\right) - \left(y_t - \hat{\mu}_t\right)\right]' \left[\left(y_t^i - \hat{\mu}_t\right) - \left(y_t - \hat{\mu}_t\right)\right]$$

$$= \frac{1}{N_t}\sum_{i=1}^{N_t}\left(y_t^i - \hat{\mu}_t\right)'\left(y_t^i - \hat{\mu}_t\right) + \frac{1}{N_t}\sum_{i=1}^{N_t}\left(y_t - \hat{\mu}_t\right)'\left(y_t - \hat{\mu}_t\right),$$

where the first term on the right-hand side is the MSE attributed to the idiosyncratic component and the second term is the MSE attributed to the common component. The cross term is zero because

$$\frac{1}{N_t}\sum_{i=1}^{N_t}\left(y_t^i - \hat{\mu}_t\right)'\left(y_t - \hat{\mu}_t\right) = \left(\hat{\mu}_t - \hat{\mu}_t\right)'\left(y_t - \hat{\mu}_t\right) = 0.$$

guide such decisions. The public know that data such as GDP are often revised and sometimes thoroughly revised. They take such unpredictable outcomes into account and make their forecasts as accurately as possible on average.

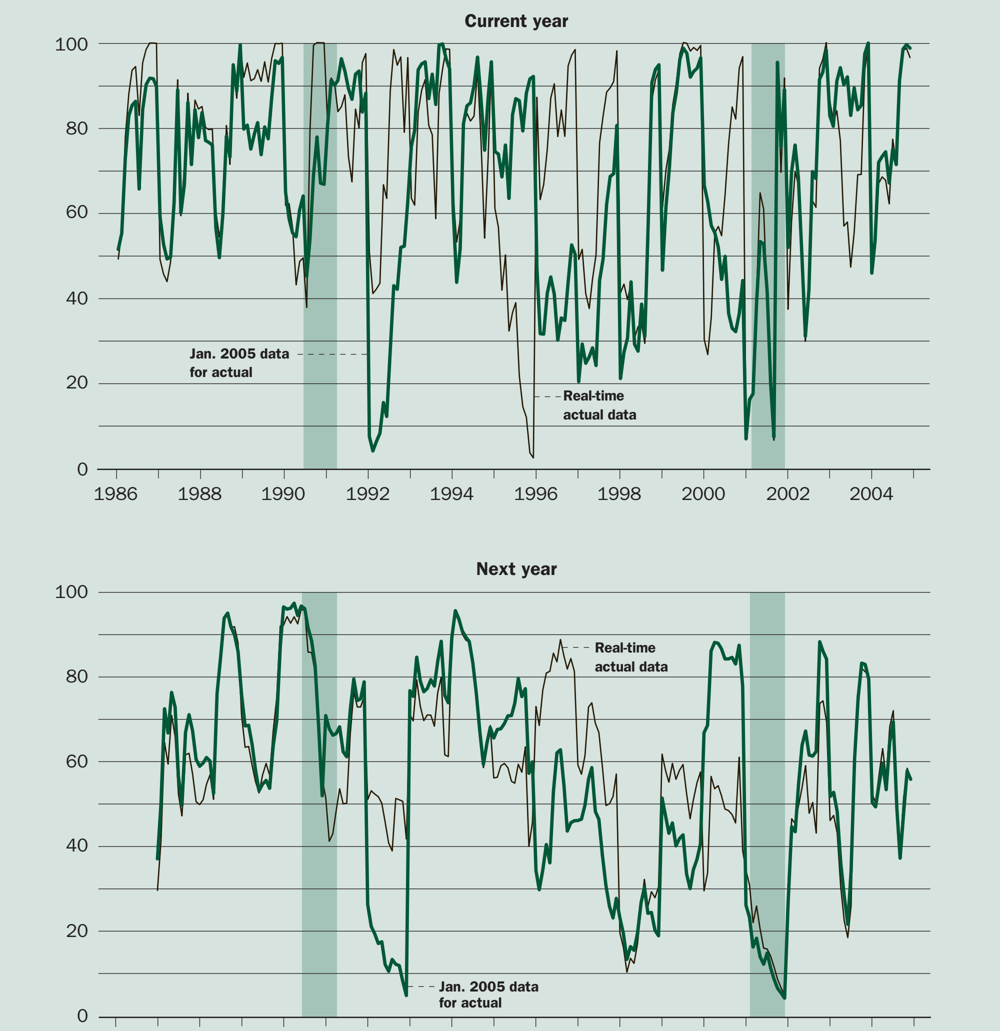
In this section, we use the revised and most current data available at the beginning of 2005 to recompute the forecast errors. Figure 15 displays the Blue Chip Consensus accuracy scores with the vintage data and the final data for both the current-year and next-year forecasts. The average current year score using vintage data is 70.9 while the average current-year score using final data is 67, just 3.9 points lower. For the next-year forecast, the average scores using vintage data and final data are very similar: 57.4 using vintage data and 56.4 using final data. During several periods (1992, 1995–96, and 1998) the next-year forecast scores are lower using final data, but several periods (1994, 1999, and 2002) have higher scores. These results indicate that future data revisions are random enough that they do not introduce a bias that significantly affects forecast scores on average. More important, the findings also suggest that the data revisions do not pose significant risks for policymakers.

One would expect, perhaps, a greater disparity between the two scores given that additional revision errors are unpredictable. However, an important advantage of using the final data is that one can avoid the distorted GDP forecast errors caused by the 1995 data revision. By comparing the first panels of Figures 6 and 11, one can see that the distortion is completely eliminated when the final data are used to measure the forecast accuracy. Still, when the 1995 period is excluded, the difference between the current-year scores using vintage and final data increases from 3.9 to 7.7. Looking more closely at the source of this difference, we find that it can be attributed mostly to the GNP/GDP forecast error.

Figure 16 displays the decompositions of forecast errors for GNP/GDP using the final data as realized values. A comparison of this figure with Figure 11 reveals some notable differences in the breakdown in the composition for both the current-year and next-year forecasts. In the first panel of Figure 11, we see larger overall errors in 1992 and in the 1996–2004 period that are due to increases in the common component

7. This interpretation is consistent with the results of Stock and Watson (2003) and Sims and Zha (2006).

Figure 15
Blue Chip Consensus Scores: Current versus Real-Time Actual Data

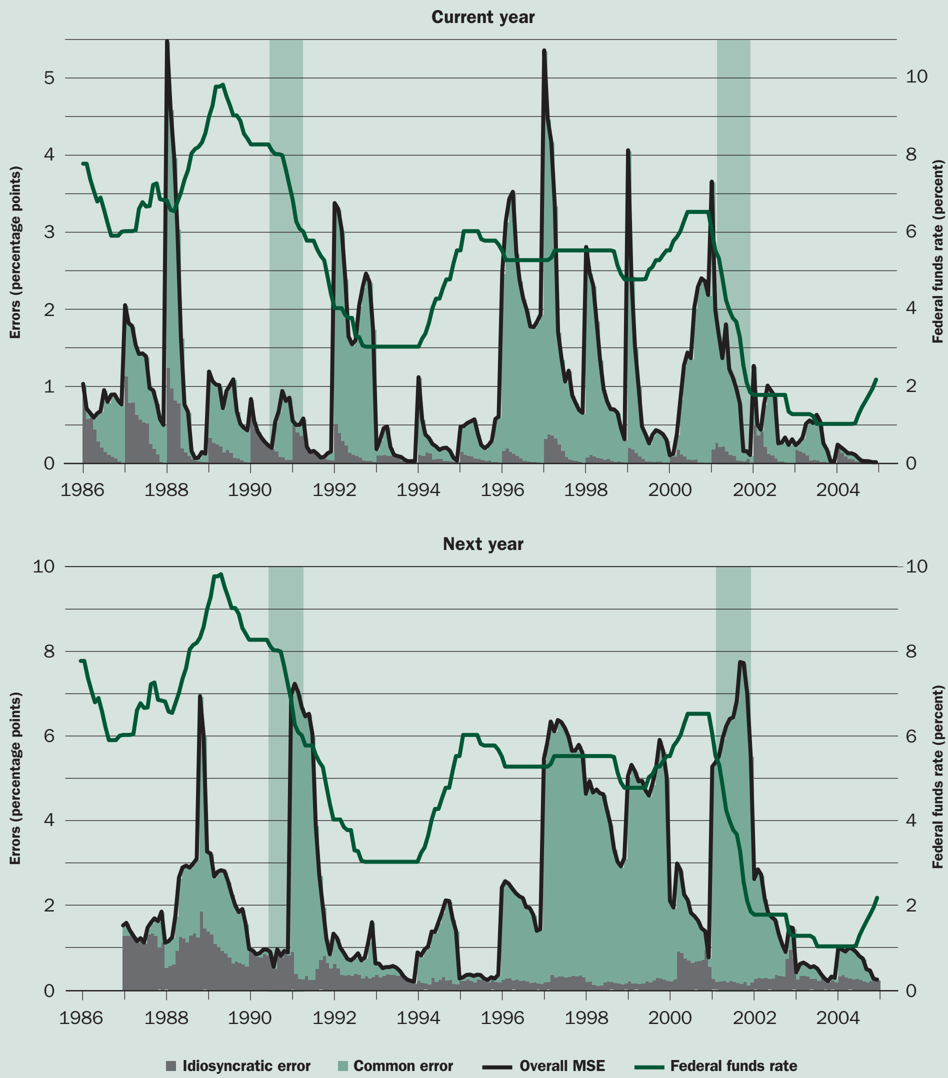


Note: The shaded vertical bars indicate recessions. In the first panel, the average score using real-time actual data is 70.9; the average score using January 2005 data for actual data is 67.0. In the second panel, the average score using real-time actual data is 57.4; the average score using January 2005 data for actual data is 56.3.

Source: Authors' calculations from monthly *Blue Chip Economic Indicators* data

of the forecast error. Consequently, a greater proportion of the error each period is due to the common component. The average contribution of the common component to the overall error rises to 73.9 percent from 56.7 percent. In addition, the overall error in 1995 using vintage data (which resulted from the changing to chain-weighted GDP) is no longer present. For the next-year forecasts in the second panel of Figure 11, we again see that the overall error has increased but to a considerably more modest degree. The overall forecast error prior to the 1990–91 recession is less using final data but is greater (on aggregate) for the 1996–2000 period. But once again, this increase

Figure 16
Mean Square Error (Using January 2005 Data as Actual Data) of GDP Forecasts



Note: The shaded vertical bars indicate recessions. In the first panel, the idiosyncratic percent of total error (per period average) is 26.1; the common percent of total error (per period average) is 73.9. In the second panel, the idiosyncratic percent of total error (per period average) is 38.2; the common percent of total error (per period average) is 61.8.

Source: Authors' calculations from monthly *Blue Chip Economic Indicators* data

in overall error is attributable to the common component. The average contribution of the common component rises to 61.8 percent from 59 percent.

Our findings suggest that using final data or vintage data may make little difference when evaluating forecasts. The results show that the average Blue Chip Consensus score is modestly affected for current-year forecasts and almost unchanged for next-year forecasts. In addition, the decrease in score for current-year and next-year forecasts results from an increase in the common component of the forecast error and does not affect the idiosyncratic component. Therefore, the effect of a switch to final data for evaluating individual forecasts scores should be roughly equal across forecasts. The use of final data eliminates the need for arbitrarily choosing among different vintages.

Conclusion

In 1994 the FOMC began to release statements after each meeting. The amount of policy information released in the statements has increased and changed over time. The findings from Kohn and Sack (2003) and Ehrmann and Fratzscher (2004) suggest that financial markets are sensitive to the information revealed in these statements. While knowing whether the statements have affected markets is important, understanding whether the statements are providing strong signals concerning the FOMC's views about the future path of the economy or economic policy is also important. That is, has the public's ability to forecast future economic and financial conditions improved since 1994? This question is important because one hopes that transparency, if appropriately communicated, enhances market participants' ability to forecast (Woodford 2005).

This article analyzes the forecast errors across a large section of forecasters and for a set of five key macroeconomic variables. The analysis finds evidence that the individuals' forecasts have been more synchronized since 1994, implying the possible effects of the FOMC's transparency. On the other hand, we find little evidence that the common forecast errors, which are the driving force of overall forecast errors, have become smaller since 1994. In fact, common forecast errors have increased and have become more volatile on several dimensions. These common errors seem to be associated with business cycles and other economic shocks. Transparent monetary policy may not necessarily enhance the public's ability to predict business cycles.

On the other hand, it is possible that we do not have a long-enough sample to observe the effects of transparency because the FOMC just began in August 2003 to provide more explicit guidance on the likely path of future policy and its contingency on future economic conditions. We hope that our findings will generate more research on this important topic.

Appendix 1
Data Description

Three-month Treasury bill rate: 1986–2004. Secondary market, monthly average. Source: Board of Governors of the Federal Reserve System.

Consumer price index: 1986–2004. CPI-U (all urban consumers). Source: U.S. Department of Labor, Bureau of Labor Statistics.

Gross national/domestic product: 1986–95, not chained; 1996–2004, chained. Source: U.S. Department of Commerce, Bureau of Economic Analysis.

Unemployment rate: 1986–2004. All workers sixteen years or older. Source: U.S. Department of Labor, Bureau of Labor Statistics.

Corporate bond yield: 1986–95. Aaa, monthly average. Source: Moody's Investors Service Inc.

Ten-year Treasury note yield: 1996–2004. Constant maturity, monthly average. Source: Board of Governors of the Federal Reserve System.

Appendix 2
Scores and Ranks for Individual Forecasters

In this appendix, the following table shows the average scores for all the individual forecasters who have continued to participate in the surveys in recent years. The table also includes the consensus forecast and the Bayesian vector autoregressive (BVAR) model. The BVAR model is often used in the empirical literature as a benchmark for model comparison (Robertson and Tallman 1999, 2001), and reporting the real-time forecasting performance of this model is of particular interest to academic researchers. For completeness, we also report other forecasters' scores toward the end of the table. The years in which each forecaster participated in the Blue Chip surveys are also reported in the table.

Table
Overall Performance: Score

	Overall		Current year		Next year		Participation	
Forecaster Name	Avg. score	Std. dev.	Avg. score	Std. dev.	Avg. score	Std. dev.	Current year	Next year
BC—average of top 10	82.24	16.86	86.45	15.99	77.81	16.66	228	216
BC—consensus	64.36	23.49	70.92	24.07	57.43	20.77	228	216
Macroeconomic Advisers, LLC	62.58	27.71	71.57	26.25	53.10	26.06	227	215
Schwab Washington Research Group	62.04	28.26	69.97	27.11	53.64	27.07	197	186
Atlanta BVAR	59.69	31.19	69.21	29.54	49.64	29.75	228	216
U.S. Trust Company	59.25	27.15	64.61	26.25	49.96	26.25	227	131
ClearView Economics	59.23	28.94	66.69	27.72	50.10	27.99	66	54
Banc of America Corporation	59.22	27.10	63.28	27.82	54.87	25.68	204	190
Northern Trust Company	58.75	28.01	63.34	27.27	53.17	27.95	222	183
Wayne Hummer & Company	55.89	27.27	58.05	27.61	53.58	26.78	228	214
Moody's Investors Service	55.04	28.03	65.77	28.63	42.35	21.34	78	66
Perna Associates	54.61	26.31	60.90	28.35	47.82	22.08	167	155

Appendix 2 (continued)

Forecaster Name	Overall Avg. score	Overall Std. dev.	Current year Avg. score	Current year Std. dev.	Next year Avg. score	Next year Std. dev.	Participation Current year	Participation Next year
Merrill Lynch	54.50	27.41	58.36	28.97	50.32	25.02	206	190
Wells Capital Management	53.58	28.71	59.83	28.19	46.83	27.81	161	149
National Association of Home Builders	53.56	26.06	58.77	26.69	47.93	24.21	176	163
Nomura Securities	52.55	28.87	55.77	29.82	48.57	27.41	63	51
National City Bank of Cleveland	52.01	26.08	56.75	26.56	46.93	24.61	224	209
DuPont	51.68	25.60	57.06	28.14	46.00	21.23	228	216
Georgia State University	51.67	27.39	51.72	28.64	51.62	26.07	223	211
Fannie Mae	51.43	28.00	59.67	29.13	41.81	23.35	84	72
DaimlerChrysler AG	51.34	29.13	58.94	29.24	43.35	26.85	226	215
Standard & Poors	51.25	30.43	58.86	30.09	42.78	28.63	120	108
Eggert Economic Enterprises	50.79	25.90	50.12	27.56	51.48	24.09	225	215
Siff, Oakley, Marks Inc.	50.66	28.19	56.56	27.41	44.77	27.78	197	197
Evans, Carrol and Associates	50.43	29.77	58.01	30.35	42.86	27.21	202	202
Bank One	49.82	31.39	56.87	31.75	42.21	29.22	205	190
Bear Stearns & Company Inc.	49.67	29.96	53.11	30.39	43.95	28.59	98	59
BC—average of individual scores	48.13	16.08	51.84	16.22	44.21	15.00	228	216
La Salle National Bank	47.47	29.73	54.13	32.16	40.22	24.97	158	145
Prudential Securities	47.07	31.41	47.40	33.06	46.57	28.88	175	117
Prudential Financial	47.01	26.68	50.54	28.97	43.31	23.55	201	192
Goldman Sachs & Company	46.28	27.19	59.47	25.49	30.49	19.85	79	66
National Association of Realtors	46.10	29.24	51.08	29.08	40.10	28.56	64	53
Conference Board	45.08	29.38	52.22	31.03	37.46	25.46	224	210
Chamber of Commerce, USA	44.97	27.68	48.35	28.34	41.20	26.50	214	192
General Motors Corporation	44.30	28.03	46.05	29.42	42.42	26.40	162	150
Econoclast	43.29	27.13	42.32	30.94	44.32	22.44	227	215
Eaton Corporation	43.04	28.51	40.92	30.07	45.37	26.62	127	115
Turning Points (Micrometrics)	43.04	27.86	41.15	29.28	45.04	26.19	185	174
Comerica	42.41	25.44	43.88	29.34	40.84	20.41	178	166
UCLA Business Forecast	42.12	30.19	45.32	32.23	38.75	27.55	227	215
Motorola Inc.	42.02	28.76	50.83	31.73	31.91	20.92	102	89
JPMorgan Chase	40.92	27.12	47.57	29.12	33.40	22.55	104	92
Kellner Economic Advisers	40.79	23.00	41.86	24.65	39.55	21.02	91	79
Genetski.com	40.46	32.50	50.61	32.88	29.53	28.37	154	143
Wachovia Securities	40.39	27.19	44.60	31.05	35.69	21.33	98	88
Federal Express Corporation	39.92	26.15	41.80	28.76	37.63	22.60	65	53
DRI-WEFA	39.02	27.09	48.32	28.48	27.99	20.65	77	65
Morgan Stanley & Company	35.95	29.39	38.27	31.92	32.30	24.75	85	54
Inforum–University of Maryland	35.72	26.46	33.15	27.15	38.46	25.47	222	208
Deutsche Banc Alex Brown	30.71	28.22	31.86	26.76	29.08	30.33	91	64
Naroff Economic Advisors	29.96	28.67	33.36	32.65	25.86	22.59	70	58
Ford Motor Company	25.80	25.12	27.32	26.09	23.69	23.73	103	74
BC—average of bottom 10	7.50	6.35	6.12	6.32	8.96	6.05	228	216

REFERENCES

Bauer, Andy, Robert A. Eisenbeis, Daniel F. Waggoner, and Tao Zha. 2003. Forecast evaluation with cross-sectional data: The Blue Chip Surveys. Federal Reserve Bank of Atlanta *Economic Review* 88, no. 2:17–31.

Ehrmann, Michael, and Marcel Fratzscher. 2004. Central bank communication: Different strategies, same effectiveness? European Central Bank, unpublished paper.

Eisenbeis, Robert A., Daniel F. Waggoner, and Tao Zha. 2002. Evaluating *Wall Street Journal* survey forecasters: A multivariate approach. *Business Economics* 37, no. 3:11–21.

Faust, Jon, and Eric M. Leeper. 2005. Forecasts and inflation reports: An evaluation. Paper presented at the Sveriges Riksbank conference "Inflation Targeting: Implementation, Communication and Effectiveness," Stockholm, June 10–12.

Kohn, Donald L., and Brian P. Sack. 2003. Central bank talk: Does it matter and why? Board of Governors of the Federal Reserve System Finance and Economics Discussion Series No. 2003-55, November.

Robertson, John C., and Ellis W. Tallman. 1999. Vector autoregressions: Forecasting and reality. Federal Reserve Bank of Atlanta *Economic Review* 84, no. 1:4–18.

———. 2001. Improving federal-funds rate forecasts in VAR models used for policy analysis. *Journal of Business and Economic Statistics* 19, no. 3:324–30.

Sims, Christopher A., and Tao Zha. 2006. Were there regime switches in U.S. monetary policy? *American Economic Review* 96, no. 1:54–81.

Stock, James H., and Mark W. Watson. 2003. Has the business cycle changed? Evidence and explanations. In *Monetary policy and uncertainty: Adapting to a changing economy*. Federal Reserve Bank of Kansas City.

Woodford, Michael. 2005. Central-bank communication and policy effectiveness. In *The Greenspan era: Lessons for the future*. Federal Reserve Bank of Kansas City.