FEDERAL RESERVE BANK OF PHILADELPHIA

Ten Independence Mall
Philadelphia, Pennsylvania 19106-1574
(215) 574-6428, www.phil.frb.org

Working Papers

Research Department

WORKING PAPER NO. 98-14

EVALUATING INFLATION FORECASTS

Dean Croushore
Federal Reserve Bank of Philadelphia

June 1998

WORKING PAPER NO. 98-14

EVALUATING INFLATION FORECASTS

Dean Croushore

Federal Reserve Bank of Philadelphia

June 1998

The views expressed in this paper are those of the author and do not necessarily represent the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. Jim Sherma provided outstanding research assistance on this project. Thanks to Sally Burke for useful editorial comments. Please send comments to the author at Research Department, Federal Reserve Bank of Philadelphia, Ten Independence Mall, Philadelphia, PA 19106-1574, or e-mail: dean.croushore@phil.frb.org.

EVALUATING INFLATION FORECASTS

ABSTRACT

In the early 1980s, economists tested inflation forecasts and found that the forecasts were very bad. Either the surveys didn't capture forecasters' expectations or forecasters didn't have rational expectations. However, the sample period being examined consisted mostly of data from the volatile 1970s, when forecasting was extremely difficult. The question is: if we run the same types of tests that were performed 15 years ago on an updated sample, will we find the same problems with the forecasts?

This paper finds that much of the empirical work from 15 years ago does not stand the test of time. The forecast errors from the surveys aren't nearly as bad today as they were in the 1970s. However, there remain some problems in the forecasts. It appears to be possible to improve inflation forecasts over some sample periods using bias regressions, and the forecasts don't pass all tests for optimality.

EVALUATING INFLATION FORECASTS

In the early 1980s, economists tested the inflation forecasts in surveys taken over the previous 20 years and found that the forecasts were terrible: they were clearly biased and they lagged behind changes in inflation. The forecast errors were far from what we'd expect if the forecasts were formed by rational economic agents. Given that this was the heyday of the rational expectations revolution, these results were quite disappointing to researchers, who concluded that either those who surveyed the forecasters weren't collecting the proper data or forecasters were irrational in their beliefs about inflation. As a result, many economists stopped paying attention to the forecast surveys.[1]

But should economists have given up on forecast surveys so easily? After all, the people being surveyed are professional forecasters who would seem to have great incentive to produce good inflation forecasts. It's hard to believe that they wouldn't report their true forecasts to the surveyors; it's equally hard to believe that they didn't have rational expectations. Now that some time has passed since the initial round of papers examining this issue, perhaps it's time to suspend disbelief and reexamine the forecasts. If we run the same types of tests on the forecasts that were performed 15 years ago, will we find the same evidence of systematic forecast errors?

To answer this question, we'll rerun the bias tests from the literature on testing inflation forecasts, then run some new experiments, to see how the inflation forecasts from the surveys hold up, given a longer sample period. Our priors are: (1) if the results of the earlier literature are confirmed, we have a continuing puzzle about why forecasters are behaving irrationally or why they aren't providing their true forecasts to the surveyors; (2) if the results of the earlier literature

---

[1]For an overview of the literature on testing survey data on expectations, see Maddala (1991).

aren't confirmed, what explains the earlier findings of bias? We begin by looking at three data sets containing inflation forecasts.

THE DATA

The first question is: what measure of inflation and forecasts of inflation should we examine? Since the data on inflation itself are very noisy (month-to-month and quarter-to-quarter movements in inflation move around substantially), we don't really care about short-term forecasts of inflation.[2] Instead, we'd like to see if forecasters are correct about the overall trend in inflation. As a result, we'll look only at inflation and forecasts of inflation over a one-year horizon.

We'll examine three different surveys of inflation forecasts: (1) the Livingston survey, which gathers forecasts from a broad group of economists, (2) the Survey of Professional Forecasters (SPF), which gathers forecasts from economists who forecast for a living, and (3) the Michigan survey, which looks at forecasts from consumers. You might expect that the Michigan group spends the least amount of resources on forecasting, while the forecasters in the SPF spend the most, with the Livingston group in the middle. In this section, we'll describe each of the surveys and examine its inflation forecasts and forecast errors.

The Livingston survey collects economists' forecasts of inflation and other economic variables twice a year and has been in existence since 1946.[3] It was started by newspaper columnist Joseph Livingston, who used the survey as source material for two columns a year

[2]See Bryan and Cecchetti (1994) for a discussion of the noise in inflation data.

[3]Croushore (1997) provides a detailed review of the survey, how it is constructed, and how it has been used in research. The Livingston survey data are available on the Web at 'http://www.phil.frb.org/econ/liv/welcome.html'.

(hence the survey's semiannual structure). In 1978, as researchers began requesting the data in greater numbers, Livingston asked the Federal Reserve Bank of Philadelphia for help in maintaining the database. The Philadelphia Fed took over the entire survey upon Livingston's death in 1989. The survey asks for forecasts of a wide variety of macroeconomic variables, including both the producer price index and consumer price index. We'll focus on the latter in this article, since we'll be comparing the forecasts to the Michigan survey, which asks about consumer prices. The Livingston questionnaire asks for forecasts of the level of the CPI (not seasonally adjusted) 12 months ahead. However, since the level of the CPI is known only with a two-month delay, the forecast is really for 14 months ahead.[4] Thus, we construct a series of the actual inflation rate over a 14-month period to compare to the survey forecast.

The Survey of Professional Forecasters may be the least well known of the three surveys, but it is likely the most accurate.[5] It was started by Victor Zarnowitz and others at the American Statistical Association and National Bureau of Economic Research in 1968 and was known as the ASA-NBER Survey. This quarterly survey asks for detailed quarterly forecasts on a wide variety of macroeconomic variables and thus requires far more effort on the part of forecasters than does the Livingston survey. The respondents are mostly those who produce forecasts as a major part of their job responsibilities in the private sector. Since the only price variable that was forecast

---

[4]Questionnaires are sent out in May and November each year, when the CPI data for April and October are known. The survey asks for forecasts of the level of the CPI in the following June and December, so they represent 14-month forecasts. However, at times in the survey's past, some of the forecasters knew the May or November CPI number, so they would be making just a 13-month forecast. We assume that doesn't happen often enough to affect the results significantly; see Carlson (1977) and Fishe (1984) for discussion of this issue.

[5]Croushore (1993) contains complete details on the survey. The survey data are available on the Web at 'http://www.phil.frb.org/econ/spf/spfpage.html'.

when the survey was begun is the GNP deflator, we'll use forecasts of that variable.[6] A difficulty with using the deflator is that it gets revised over time; as a result, we use as the actual value the number recorded in the national income accounts in the third month after the end of each quarter; this prevents revisions from distorting the numbers we use as realized values in a way that forecasters could not be expected to foresee.

The Michigan survey of consumers has collected data on expectations of inflation for over 50 years, with a sample size generally above 500 respondents.[7] However, the data have been collected quarterly (in the middle month of the quarter) only since the second quarter of 1968. As part of a much larger questionnaire, the survey asks respondents, "By about what percent do you expect prices to go up, on the average, during the next 12 months?"[8] After 1977, the survey was taken monthly; to be consistent with data from previous years, we use the middle month of the quarter. Since the survey doesn't specify a particular price index but questions consumers whose knowledge of economic data is likely to be limited, we compare the survey forecasts to the Consumer Price Index (CPI), which is the most widely known measure of prices.

---

[6]The variable itself did change both in 1992, when the national income accounts switched from GNP to GDP, and again in 1996, when the GDP price index became the key price variable. However, the GNP deflator, GDP deflator, and GDP price index behave quite similarly, and there is no apparent break in the series in either 1992 or 1996.

Unfortunately, the Survey of Professional Forecasters didn't begin asking for forecasts of the CPI inflation rate until 1981, so we use deflator forecasts. Thus results in this paper won't be directly comparable to those from the Livingston and Michigan surveys, which use the CPI.

[7]See Noble and Fields (1982) for a detailed description of the survey.

[8]This quantitative question was asked beginning in the third quarter of 1977. Prior to that, respondents were asked to report a range in which they expected prices to rise, for example, "prices will rise 2 to 4 percent." For the period from 1968Q2 to 1977Q2, the method developed by Juster and Comment (1980) is used to calculate the mean expected increase in prices.

To help in interpreting the formal statistical tests on the forecasts that will be shown later, let's first look at some basic plots of the data. Figure 1 shows the data from the Livingston survey, from 1954H1 to 1996H2, where both the one-year-ahead forecast and the actual value are shown in the upper panel.[9] In the lower panel, the forecast error is shown (defined as the actual value minus the forecast). Figures 2 and 3 show similar plots for the Michigan survey and the Survey of Professional Forecasters.[10] All three figures show the same basic patterns: (1) forecasters didn't think inflation would rise as high as it did as a result of the oil price increases in the 1970s; (2) forecasters were surprised at how low inflation was in the early 1980s; and (3) forecasters have been forecasting an upturn in inflation in the 1990s that hasn't happened. Despite these obvious forecast errors (which must not have been so obvious to the forecasters at the time), we want to know if statistical tests indicate that the forecasts are poor.

BIAS TESTS FROM THE EARLY 1980S

One of the first tests of inflation forecasts was a test for *bias*. A set of inflation forecasts over time is biased if a regression of the actual values for inflation (the dependent variable) on a

---

[9]Note that the dates on the horizontal axis represent the dates at which a forecast was made. The corresponding "forecast" point is the forecast for the 14-month period (April to June or October to December) that was made at that date, while the "actual" data point is the actual value of inflation over that period. So the data points shown for 1996H2 are: (1) the forecast from the December 1996 Livingston survey for the CPI inflation rate from October 1996 to December 1997; and (2) the actual inflation rate in the CPI from October 1996 to December 1997.

[10]For the SPF, the data points are the forecast for the GNP (GDP) deflator (price index) from the current quarter to four quarters ahead and the actual value over that period. For the Michigan survey, the data points are the forecast for the CPI inflation rate over the coming year to the actual value over that period. For example, for 1996Q4, the forecast is the November 1996 Michigan survey for the inflation rate over the next 12 months and the actual inflation rate in the CPI from November 1996 to November 1997.

constant term and the forecasted values (the independent variable) yields coefficients that are significantly different from 0 for the constant term and 1 for the forecast term. That is, the regression is:

$$\pi_t = \alpha + \beta \pi_t^f + \varepsilon_t, \tag{1}$$

where π_t is the actual inflation rate and π_t^f is the forecast at each date t. The bias test is simple and sensible: over a long sample period, you'd expect $\hat{\alpha}$ to be close to zero and $\hat{\beta}$ to be close to one. If not, you could improve on the forecasts in the following way. If you ran the regression from equation (1) and found estimates of $\hat{\alpha}$ and $\hat{\beta}$, you could form a new forecast, π_t^i, where *i* stands for "improved," by calculating:

$$\pi_t^i = \hat{\alpha} + \hat{\beta}\pi_t^f. \tag{2}$$

Theil (1966) was perhaps the first researcher to use this methodology both for analyzing possible bias in forecasts, as in (1), and for improving the forecasts, as in (2). There is, however, some question about what it means if $\hat{\alpha} \neq 0$ or $\hat{\beta} \neq 1$. As Webb (1987) points out, just because the regression in (1) shows bias doesn't mean that forecasts are lousy or can be improved by (2), because there are revisions to data, because the coefficients of (1) may change over time, or because forecasters may think there's some probability that a major policy change will occur, which could lead to persistent forecast errors, such as the peso problem. So we'll need to look for evidence of these problems as we go along.

Some very early studies, such as Turnovsky (1970), showed that inflation forecasts from the Livingston survey were biased using the test in (1). Turnovsky's original results are shown in

Table 1. He studied both the periods from 1954 to 1964 (line 1 in Table 1) and 1962 to 1969 (line 2), finding bias in both periods. However, later research suggested two problems with Turnovsky's analysis: (1) he used bad data; and (2) he didn't account properly for overlapping observations. The first point, that Turnovsky used bad data, was made by Carlson (1977), who discovered that newspaper columnist Joseph Livingston, who ran the survey, was modifying the forecasts to reflect data that were released between the time the survey was taken and when the results were published. Carlson was able to get the original data and create the correct data set, which he then used. The second point, that Turnovsky didn't account properly for overlapping observations, arises because the forecasts covered a 14-month period, but a forecast was produced every six months. Since an inflation shock would thus affect two or three consecutive forecast errors, any regression might show signs of autocorrelation. To conduct the bias test properly, a researcher must either construct a data set of nonoverlapping observations (in this case, by just using one of every three observations) or modify the covariance matrix as suggested by Brown and Maital (1981), using the method of Hansen and Hodrick (1980), perhaps as modified by Newey and West (1987), to guarantee a positive definite covariance matrix.

When we rerun Turnovsky's tests using the corrected data and Newey-West's methods, we find support for Turnovsky's suggestion that the Livingston survey is biased in both of his sample periods (lines 3 and 4 of Table 1).[11] The last column is the significance level of a chi-square test of the null hypothesis that " = 0 and $ =1; the null is rejected in both sample periods that are replicated. But if we run the test for the entire sample period, through the fourth quarter

---

[11]We don't use the method of splitting the sample into three parts to eliminate the problem of overlapping observations because the sample period is so short.

of 1996, we find no evidence of bias at all (line 5 of Table 1). The same is true when using nonoverlapping observations.[12] Thus, Turnovsky's results on the bias of the inflation forecasts don't hold up over the entire sample.

Brown and Maital (1981) were the first to suggest the use of the Hansen-Hodrick method. Their results, shown on line 1 of Table 2, indicate no bias in the forecasts, with a p-value of the P^2 test of .195. However, our replication of their results, shown in line 2, shows bias, as the p-value is just .005. Brown and Maital evidently used the uncorrected Livingston data, even though they were aware of the problems that Carlson had pointed out. Thus, the results of our replication differ substantially from the original, as they did for Turnovsky as well. When we update the data through 1996, once again all evidence of bias disappears, as the p-value rises to .446. Similar results obtain (not shown) using the nonoverlapping observations data.

Bryan and Gavin (1986) also tested the Livingston survey inflation forecasts for bias, finding evidence of bias using both the Hansen-Hodrick procedure and the nonoverlapping observations methods.[13] Line 1 of Table 3 shows their results using the Hansen-Hodrick method. We are able to replicate their results fairly closely, as shown in line 2. But when we update the data through the end of 1996, we no longer reject the null hypothesis that " = 0 and $ =1, with

---

[12]The results on the nonoverlapping observations are too numerous to show here, but the p-values on the chi-square test for " = 0 and $ =1 are in all cases greater than .10, so we wouldn't reject the null hypothesis.

[13]Bryan and Gavin also test the Michigan survey but use data from the early period of the survey, which asked only if people thought prices would rise or fall, rather than asking them to give a point estimate. The assumptions used (by earlier researchers at the University of Michigan) to turn the directional data into point estimates seem dubious.

the p-value for the chi-squared test at .184. So, again, although the early data suggested that the forecasts were biased, the addition of more data suggests that the forecasts are not biased.

What about the Survey of Professional Forecasters and the Michigan survey of consumers? Since data aren't available for those surveys until 1968, we can't run the bias test regression over the same sample period as for the Livingston survey. But if we cut the sample into two parts, 1968 to 1983 and 1983 to 1996, we can see results very similar to those of the Livingston survey. As shown in Table 4, running the bias test for the SPF leads to a finding of bias in the first subsample but no bias in the data that extend through 1996, as was the case for the Livingston survey results. For the Livingston survey, Table 5 shows no bias in the data in either the subsample from 68Q2 to 83Q4 or in the full sample ending in 96Q4.

An even clearer way to see the lack of bias in the forecasts is to examine scatterplots of the forecasts and actual values, as shown in Figure 4 for the Livingston survey, Figure 5 for the SPF, and Figure 6 for the Michigan survey. In all three figures, a 45-degree line has been drawn in. If the forecasts were biased, they would not lie along the 45-degree line but would differ from it in some systematic fashion. No bias is apparent in any of the figures, a result confirmed by the regressions in Tables 1, 2, and 3. However, you can see the effect of the oil-price shocks of the 1970s in the figures; they are the points in the scatterplots where actual inflation is a lot higher than the forecast of inflation. Indeed, had we shown the scatterplots prior to 1983, most of the data points would have been above the 45-degree line. That's why the tests run in the 1970s and early 1980s showed bias in the surveys, while the tests run today do not. The question is: knowing this pattern of forecast errors, could someone have used the knowledge to make better forecasts?

FORECAST-IMPROVEMENT EXERCISES

Examining the pattern of the forecast errors over time might lead one to think that, even though the bias tests don't show bias over the entire sample period, one could run the bias regression over different subsamples of the data, modify the forecasts appropriately, and produce a better forecast. That turns out to be very difficult to do for several reasons. First, the figures showing the forecast errors are a bit misleading because of the overlapping observations problem. That is, a shock (such as a sharp rise in oil prices) induces forecast errors in four consecutive observations for the SPF and Michigan surveys and two consecutive observations for the Livingston survey. Second, although it's obvious after the fact that the forecast errors have switched from being consistently positive in the 1970s to frequently negative in the 1990s, figuring out when that switch occurred in real time isn't easy. Third, with inflation forecasts covering one year ahead, you don't know if you've made a forecast error for over a year (since there's a year for the forecast horizon and a lag in learning the final data).

To illustrate the problems in improving the survey forecasts, we run the following exercise to try to improve the forecasts. We use a set of rolling regressions to estimate equation (1), using different data samples, then modify the survey forecast as described in equation (2). We then calculate the root-mean-square-forecast error, to see if the "improved" forecast is better than the survey forecast. We do this for three different sets of regressions, first, using the entire data sample up to that date as the regression "window"; second, using 10 years of inflation and survey data as the regression window; and, third, using five years of inflation and survey data as the regression window. So, for example, with the SPF data, we regress the actual value of inflation from 1968Q4 to 1969Q4 on the one-year-ahead inflation forecast that was made in 1968Q4, recognizing that the actual value for inflation isn't available until 1970Q1. We use that regression to form an improved one-year-ahead inflation forecast for 1970Q1. Then we advance the data through 1970Q1 and repeat the exercise for the 1970Q2 forecast, continuing this set of rolling regressions through the end of the data sample in 1996Q4. Shorter data windows (like the five-year window as opposed to the 10-year window or the full-sample window) allow for quicker response to structural change in the forecast errors but may be less stable over time.

The results of these exercises are shown in Table 6. We begin by running the exercise using the first available period over which we have at least five years of data on the forecasts and actuals. For the Livingston survey, this is 1959H1; for the SPF it is 1973Q4; and for the Michigan survey it is 1973Q2. The results are shown in the first three rows of the table. It is clear that over this full period, attempts to reduce the root-mean-square-forecast error by using these methods fail, as the RMSFEs of the modified forecasts are all higher than for the original survey forecast.

We next investigate whether this method works over subsamples, looking at samples from 1974 to 1996, 1980 to 1996, 1985 to 1996, and 1990 to 1996. The surveys can't be improved on for the sample from 1974 to 1996. For the period from 1980 to 1996, the Michigan survey can be improved on, using a five-year window of data. For the period from 1985 to 1996, any of the three choices of data windows leads to a reduction in RMSFE for the Michigan survey. Clearly, the Michigan survey is most amenable to forecast improvement by accounting for bias. Finally, for the period from 1990 to 1996, it's possible to improve on all three surveys, using any of the three choices of data windows for the Michigan survey, or using the five-year or 10-year windows for the SPF and the Livingston survey. However, the reduction in RMSFE is not very large for the latter two surveys.

A look at figures 1, 2, and 3 suggests why this forecast improvement is possible. Because the forecast errors have persisted in one direction for extended periods, it is possible to improve on the forecasts a little bit. But there's enough variation in the forecast errors that it's hard to improve on the forecasts by very much. In addition, any sample period that includes the time when the forecast errors switched from consistently positive to consistently negative (around 1980) is one in which it is hard to improve the forecasts, because the forecast errors get even larger around that time.

OTHER TESTS FOR ACCURATE FORECASTS

Bias tests aren't the only way to evaluate inflation forecasts, though they have received the most attention in the literature. There are a variety of tests that good forecasts should pass. This section reports on tests suggested by Diebold and Lopez (1996).

Sign Test. The first test we'll examine is the sign test. If a forecast is optimal, the forecast errors should be independent with a zero median. The sign test examines this null hypothesis by examining the number of positive observations in the sample, which has a binomial distribution. The sign test requires that the observations be nonoverlapping, so we cut the sample into three parts for the Livingston survey and four parts for the SPF and Michigan survey.

The results are shown in Table 7. The top half of the table shows results for the longest sample period possible (which differs across the surveys), while the bottom half shows results for the period from the end of 1968 to the end of 1996 (the same period for all three surveys). We begin by running the sign test over the entire sample (rows with subsamples labeled "Full"), though the observations aren't independent, so it isn't surprising that we reject the null hypothesis for the SPF and the Michigan survey. When we cut the samples into subsamples, we don't reject the null for any of the subsamples over any sample period, except for the second Livingston subsample over the full sample period. Choosing a significance level of .05, we reject the null for a sample split into *n* subsamples at a significance level of .05/*n*. Thus, we don't reject the hypothesis of independent forecast errors with a zero median in 10 of the 11 subsamples; the one rejection, however, does suggest that the Livingston survey forecast errors aren't independent. Over the 1968 to 1996 period, shown in the lower half of the table, we don't reject the null hypothesis for any of the subsamples.

Wilcoxon Signed-Rank Test. The Wilcoxon signed-rank test is related to the sign test, as it has the same null hypothesis, but requires distributional symmetry. It accounts for the relative sizes of the forecast errors, not just their sign. The test statistic is the sum of the ranks of the absolute values of the positive forecast errors, where the forecast errors are ranked in increasing order.

The results of the test are shown in Table 8. As with the sign test, the top half of the table shows results for the longest sample period possible, while the bottom half shows results for the period from 1968 to 1996. The null hypothesis is never rejected for any of the independent subsamples in any of the surveys in either of the sample periods, suggesting that the forecast errors are independent and symmetrically distributed with zero median.

Zero-Mean Test. A simple test that optimal forecasts should pass is that the mean of the forecast errors should be zero. Again, however, because of overlapping observations, this won't be a valid test for the full sample of forecasts, but splitting the sample into parts (three subsamples for Livingston, four for SPF and Michigan) will provide a sensible test. As in the preceding tables, the top half of the table shows results for the longest sample period possible, while the bottom half shows results for the period from 1968 to 1996.

As Table 9 shows, all the surveys pass the zero-mean test with flying colors. The only rejection of the null hypothesis is for the Livingston survey's full sample. But that rejection doesn't mean much because of the overlapping observations. Thus the surveys have forecast errors with zero means.

Dufour Test. Dufour adapts the Wilcoxon signed-rank test and applies it to the product of successive forecast errors. This is a stringent test of the hypothesis that the forecasts errors are white noise, in particular that they are symmetric about zero.

A look at figures 1, 2, and 3 of the forecast errors suggests that the surveys are unlikely to pass this test, since there's some obvious serial correlation in the forecast errors. But those figures are misleading, once again, because of the overlapping observations, which induce serial correlation. By splitting the sample into independent subsamples, however, we can test them for serial dependence, with the results shown in Table 10. In each of the surveys, in both of the sample periods, at least one subsample rejects the null hypothesis of serial independence, so all three surveys fail the test. Thus the forecast errors are not white noise.

Pearce Test. Pearce (1979) ran a very simple and convincing test that the forecasts failed. He took the actual values for the monthly CPI, estimated an ARIMA model on them, and used the model to forecast the future inflation rate. The model was simple (an IMA(1,1) process on inflation) and used only data on inflation itself, yet Pearce showed that these forecasts were far superior to the Livingston survey data, based on the mean squared forecast errors. The first two lines of Table 11 show Pearce's original results and our attempt to replicate those results on data from 1959H1 to 1975H2.[14] We are able to replicate Pearce's work fairly closely, though there may be some differences in the results because of the precise definition of the CPI he used (seasonally adjusted or not) and the estimation of the ARIMA model. But Pearce's main result is simply that the ARIMA model has a lower mean-squared error than the Livingston survey. Thus, the survey yields poor forecasts, since it can be bettered by a simple univariate time-series model.

However, as line (3) of the table shows, the ARIMA model does worse in the 20 years following Pearce's sample. Over that period, the ARIMA model's mean-squared error is 4.54, compared to the Livingston survey's 3.65. When we combine the two subsamples in line (4), we find that there isn't much difference between the mean-squared errors over the entire period.

These results underscore the fact that one's evaluation of the quality of survey forecasts of inflation depends strongly on the sample period. You can find periods when forecasts looked bad, like the 1970s, and other times when they looked good, like the 1980s. So economists were perhaps premature in thinking the survey forecasts were worthless, but at the same time, the forecasts don't inspire a tremendous amount of confidence, either.

CONCLUSION

Surveys of inflation forecasts have had a bad reputation. On the basis of statistical tests in the early 1980s, economists had doubts about how accurate the forecasts were. But that was largely the effect of the oil-price shocks in the 1970s. If we look at the data today, the forecasts

[14]Note that Pearce uses a different timing convention; he uses the date at which the forecast is realized while we use the date at which a forecast is made. So his "1960(6)" date is the same as our "1959H1" date.

look much better. Nonetheless, it appears possible to improve inflation forecasts over some sample periods using bias regressions, and the forecasts don't pass all tests for optimality.

Table 1

Turnovsky Replication

	Sample period	$\hat{\alpha}$	$\hat{\beta}$	$\bar{R}^2$	D.W.	χ^2
Turnovsky's results:						
(1)	54H1-64H2	1.335 (0.62)	0.347 (0.39)	0.55	1.59	NA
(2)	62H1-69H2	--	1.272 (0.105)	0.70	1.93	NA
Replication with corrected data & improved estimator:						
(3)	54H1-64H2	0.889 (0.422)	0.798 (0.385)	0.08	0.61	.044
(4)	62H1-69H2	-0.316 (0.320)	1.796 (0.134)	0.89	1.49	.000
(5)	54H1-96H2	0.549 (0.424)	1.019 (0.150)	0.62	0.41	.109

Note: standard errors are in parentheses. NA = not available.

Table 2
Brown-Maital Replication

	Sample period	$\hat{\alpha}$	$\hat{\beta}$	$\bar{R}^2$	D.W.	χ^2
Brown-Maital's results:						
(1)	61H2-77H2	0.97	1.01	NA	NA	.195
		(1.07)	(0.25)			
Replication with corrected data:						
(2)	61H2-77H2	0.820	1.181	0.64	0.57	.005
		(0.515)	(0.199)			
Update:						
(3)	61H2-96H2	0.219	1.079	0.56	0.40	.446
		(0.709)	(0.201)			

Note: standard errors are in parentheses. NA = not available.

Table 3
Bryan-Gavin Replication

	Sample period	$\hat{\alpha}$	$\hat{\beta}$	$\bar{R}^2$	D.W.	χ^2
Bryan-Gavin's results:						
(1)	56H1-79H2	0.598	1.321	0.77	0.61	.000
		(0.6)	(0.15)			
Replication:						
(2)	56H1-79H2	0.571	1.321	0.79	0.64	.000
		(0.406)	(0.150)			
Update:						
(3)	56H1-96H2	0.521	1.024	0.60	0.40	.184
		(0.524)	(0.176)			

Note: standard errors are in parentheses. NA = not available.

Table 4
Results for Survey of Professional Forecasters

	Sample period	‟	$̂	$\bar{R}^2$	D.W.	P^2
First subsample:						
(1)	68Q4-83Q4	3.211 (1.245)	0.565 (0.187)	0.21	0.17	.036
Full sample:						
(2)	68Q4-96Q4	0.430 (0.820)	0.935 (0.158)	0.48	0.19	.867

Note: standard errors are in parentheses. NA = not available.

Table 5
Results for Michigan Survey

	Sample period	α	$\hat{\beta}$	$\bar{R}^2$	D.W.	χ^2
First subsample:						
(1)	68Q2-83Q4	0.125	1.061	0.55	0.69	.541
		(1.400)	(0.200)			
Full sample:						
(2)	68Q2-96Q4	-1.512	1.235	0.64	0.79	.236
		(0.895)	(0.154)			

Note: standard errors are in parentheses. NA = not available.

Table 6
RMSFEs for Forecast-Improvement Exercises

Survey	Period	Original Survey	Attempts to Improve on Survey: Full Sample	10-year Window	5-year Window
Livingston	‘59H1-’96H2	2.02	2.25	2.36	2.90
SPF	‘73Q4-’96Q4	1.54	2.66	2.58	2.62
Michigan	‘73Q2-’96Q4	1.88	2.39	2.38	2.32
Livingston	‘74H1-’96H2	1.89	2.46	2.60	3.28
SPF	‘74Q1-’96Q4	1.40	2.50	2.41	2.45
Michigan	‘74Q1-’96Q4	1.71	2.13	2.12	2.06
Livingston	‘80H1-’96H2	1.26	2.48	2.46	2.81
SPF	‘80Q1-’96Q4	1.10	1.86	1.70	1.60
Michigan	‘80Q1-’96Q4	1.58	1.74	1.72	1.52
Livingston	‘85H1-’96H2	1.00	1.65	1.16	1.22
SPF	‘85Q1-’96Q4	0.78	1.51	0.81	0.80
Michigan	‘85Q1-’96Q4	1.51	1.47	1.35	1.36
Livingston	‘90H1-’96H2	0.75	1.45	0.66	0.70
SPF	‘90Q1-’96Q4	0.73	1.33	0.57	0.54
Michigan	‘90Q1-’96Q4	1.59	1.43	1.03	1.13

Table 7
Sign Test

Survey	Period	Subsample	N	Reject null?	p-value
Livingston	‘54H1-’96H2	Full	86	yes	.03
		1	29	no	.58
		2	28	yes	.01
		3	29	no	.58
SPF	‘68Q4-’96Q4	Full	113	yes	.01
		1	28	no	.26
		2	28	no	.26
		3	28	no	.06
		4	29	no	.35
Michigan	‘68Q2-’96Q4	Full	115	yes	.03
		1	28	no	.13
		2	29	no	.58
		3	29	no	.19
		4	29	no	.35

Results over same sample period:

Survey	Period	Subsample	N	Reject null?	p-value
Livingston	‘68H2-’96H2	Full	57	no	.51
		1	19	no	.11
		2	19	no	.25
		3	19	no	.49
SPF	‘68Q4-’96Q4	Full	113	yes	.01
		1	28	no	.26
		2	28	no	.26
		3	28	no	.06
		4	29	no	.35
Michigan	‘68Q4-’96Q4	Full	113	yes	.02
		1	28	no	.13
		2	28	no	.45
		3	28	no	.13
		4	29	no	.35

Table 8
Wilcoxon Signed-Rank Test

Survey	Period	Subsample	N	Reject null?	p-value
Livingston	‘54H1-’96H2	Full	86	yes	.02
		1	29	no	.29
		2	28	no	.03
		3	29	no	.27
SPF	‘68Q4-’96Q4	Full	113	no	.40
		1	28	no	.80
		2	28	no	.84
		3	28	no	.44
		4	29	no	.64
Michigan	‘68Q2-’96Q4	Full	115	no	.08
		1	28	no	.21
		2	29	no	.67
		3	29	no	.22
		4	29	no	.36

Results over same sample period:

Survey	Period	Subsample	N	Reject null?	p-value
Livingston	‘68H2-’96H2	Full	57	no	.72
		1	19	no	.69
		2	19	no	.30
		3	19	no	.87
SPF	‘68Q4-’96Q4	Full	113	no	.40
		1	28	no	.80
		2	28	no	.84
		3	28	no	.44
		4	29	no	.64
Michigan	‘68Q4-’96Q4	Full	113	no	.07
		1	28	no	.21
		2	28	no	.58
		3	28	no	.21
		4	29	no	.36

Table 9
Zero-Mean Test

Survey	Period	Subsample	N	Reject null?	p-value
Livingston	‘54H1-’96H2	Full	86	yes	.00
		1	29	no	.11
		2	28	no	.05
		3	29	no	.15
SPF	‘68Q4-’96Q4	Full	113	no	.46
		1	28	no	.55
		2	28	no	.59
		3	28	no	.96
		4	29	no	.81
Michigan	‘68Q2-’96Q4	Full	115	no	.29
		1	28	no	.44
		2	29	no	.92
		3	29	no	.41
		4	29	no	.64

Results over same sample period:

Survey	Period	Subsample	N	Reject null?	p-value
Livingston	‘68H2-’96H2	Full	57	no	.12
		1	19	no	.58
		2	19	no	.19
		3	19	no	.46
SPF	‘68Q4-’96Q4	Full	113	no	.46
		1	28	no	.55
		2	28	no	.59
		3	28	no	.96
		4	29	no	.81
Michigan	‘68Q4-’96Q4	Full	113	no	.26
		1	28	no	.44
		2	28	no	.85
		3	28	no	.37
		4	29	no	.64

Table 10
Dufour Test

Survey	Period	Subsample	N	Reject null?	p-value
Livingston	‘54H1-’96H2	Full	85	yes	.00
		1	28	no	.04
		2	27	yes	.00
		3	28	yes	.00
SPF	‘68Q4-’96Q4	Full	112	yes	.00
		1	27	no	.02
		2	27	no	.02
		3	27	yes	.01
		4	28	no	.02
Michigan	‘68Q2-’96Q4	Full	114	yes	.00
		1	27	no	.05
		2	28	no	.16
		3	28	yes	.01
		4	28	no	.11

Results over same sample period:

Survey	Period	Subsample	N	Reject null?	p-value
Livingston	‘68H2-’96H2	Full	56	yes	.00
		1	18	no	.53
		2	18	no	.03
		3	18	yes	.01
SPF	‘68Q4-’96Q4	Full	112	yes	.00
		1	27	no	.02
		2	27	no	.02
		3	27	yes	.01
		4	28	no	.02
Michigan	‘68Q4-’96Q4	Full	112	yes	.00
		1	27	no	.05
		2	27	no	.18
		3	27	yes	.01
		4	28	no	.11

Table 11
Pearce Replication

	Sample period	Livingston MSE	IMA(1, 1) MSE
Pearce's results:			
(1)	59H1-75H2	4.56	3.09
Replication:			
(2)	59H1-75H2	4.57	3.33
Sample since 1976:			
(3)	76H1-96H2	3.65	4.54
Overall sample:			
(4)	59H1-96H2	4.06	4.00

Note: standard errors are in parentheses. NA = not available.

REFERENCES

Brown, Bryan W., and Shlomo Maital. "What Do Economists Know? An Empirical Study of Experts' Expectations," *Econometrica* 49 (March 1981), pp. 491-504.

Bryan, Michael F., and Stephen G. Cecchetti, "Measuring Core Inflation," in N. Gregory Mankiw, ed., *Monetary Policy*. Chicago: University of Chicago Press, 1994, pp. 195-215.

Bryan, Michael F., and William T. Gavin. "Models of Inflation Expectations Formation: A Comparison of Household and Economist Forecasts," *Journal of Money, Credit, and Banking* 18 (November 1986), pp. 539-43.

Carlson, John A. "A Study of Price Forecasts," *Annals of Economic and Social Measurement* 6 (Winter 1977), pp. 27-56.

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

Croushore, Dean. "The Livingston Survey: Still Useful After All These Years," Federal Reserve Bank of Philadelphia *Business Review*, March/April 1997, pp. 15-27.

Diebold, Francis X., and Jose A. Lopez. "Forecast Evaluation and Combination," in G.S. Maddala and C.R. Rao, eds., *Handbook of Statistics*. Amsterdam: North Holland, 1996, pp. 241-68.

Fishe, Raymond P.H. "On Testing Hypotheses Using the Livingston Price Expectations Data," *Journal of Money, Credit and Banking* 16 (November 1984, part 1), pp. 520-7.

Hansen, Lars-Peter, and Robert J. Hodrick. "Foreign Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy* 88 (October 1980), pp. 829-53.

Juster, F. Thomas, and Robert Comment. "A Note on the Measurement of Price Expectations," Institute for Social Research working paper, University of Michigan (1980).

Maddala, G.S. "Survey Data on Expectations: What Have We Learnt?" in Marc Nerlove, ed., *Issues in Contemporary Economics, vol. II. Aspects of Macroeconomics and Econometrics.* New York: New York University Press, 1991.

Newey, Whitney K., and Kenneth D. West. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55 (May 1987), pp. 703-8.

Noble, Nicholas R., and T. Windsor Fields. "Testing the Rationality of Inflation Expectations Derived from Survey Data: A Structure-Based Approach," *Southern Economic Journal* 49 (October 1982), pp. 361-73.

Pearce, Douglas K. "Comparing Survey and Rational Measures of Expected Inflation," *Journal of Money, Credit and Banking* 11 (November 1979), pp. 447-56.

Taylor, Herb. "The Livingston Surveys: A History of Hopes and Fears," Federal Reserve Bank of Philadelphia *Business Review* (January/February 1992), pp. 15-27.

Theil, Henri. *Applied Economic Forecasting*. Amsterdam: North Holland, 1966.

Turnovsky, Stephen J. "Empirical Evidence on the Formation of Price Expectations," *Journal of the American Statistical Association* 65 (December 1970), pp. 1441-54.

Webb, Roy H. "The Irrelevance of Tests for Bias in Series of Macroeconomic Forecasts," Federal Reserve Bank of Richmond *Economic Review* (November/December 1987), pp. 3-9.

Figure 1
Livingston Survey

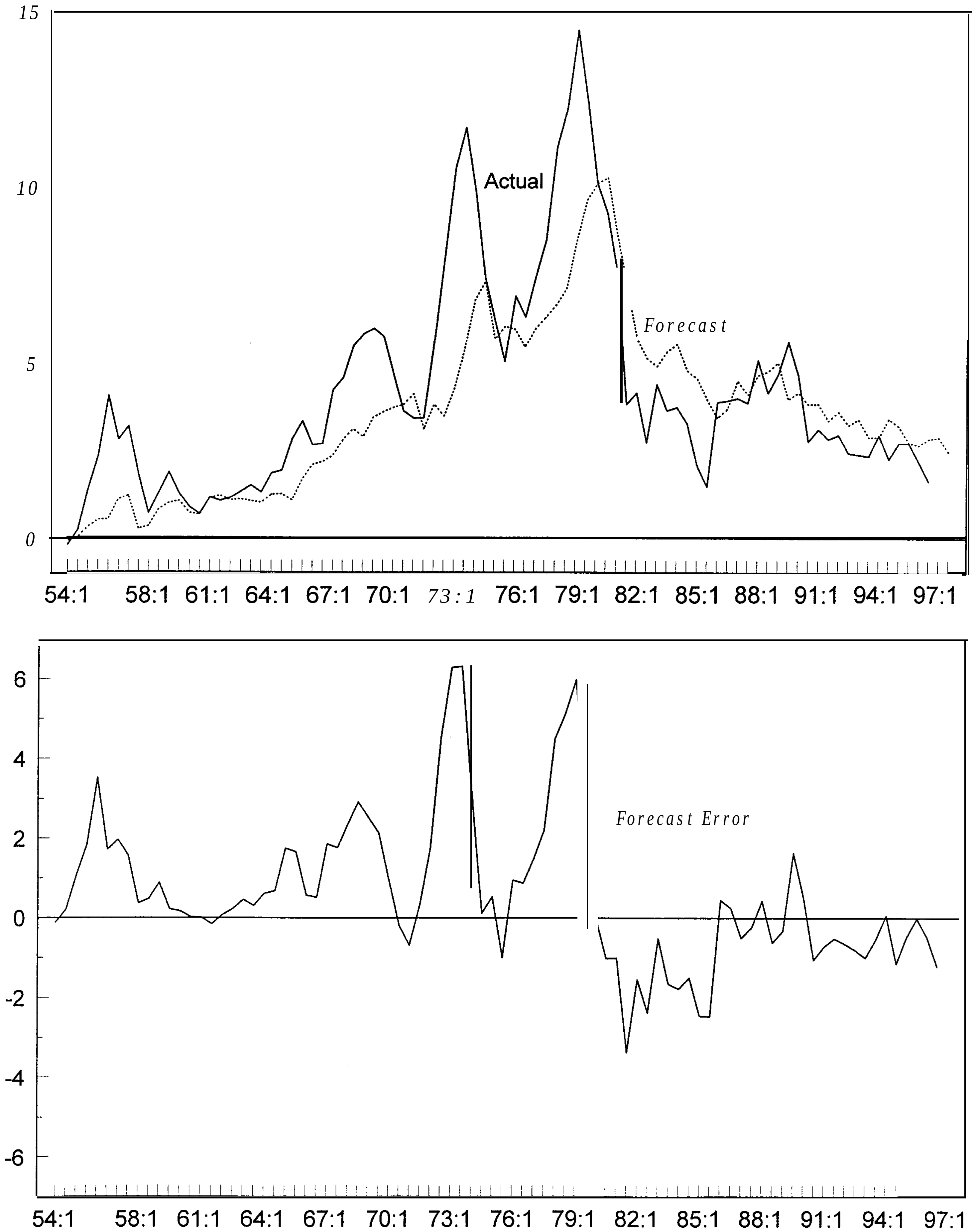


Figure 2
Survey of Professional Forecasters

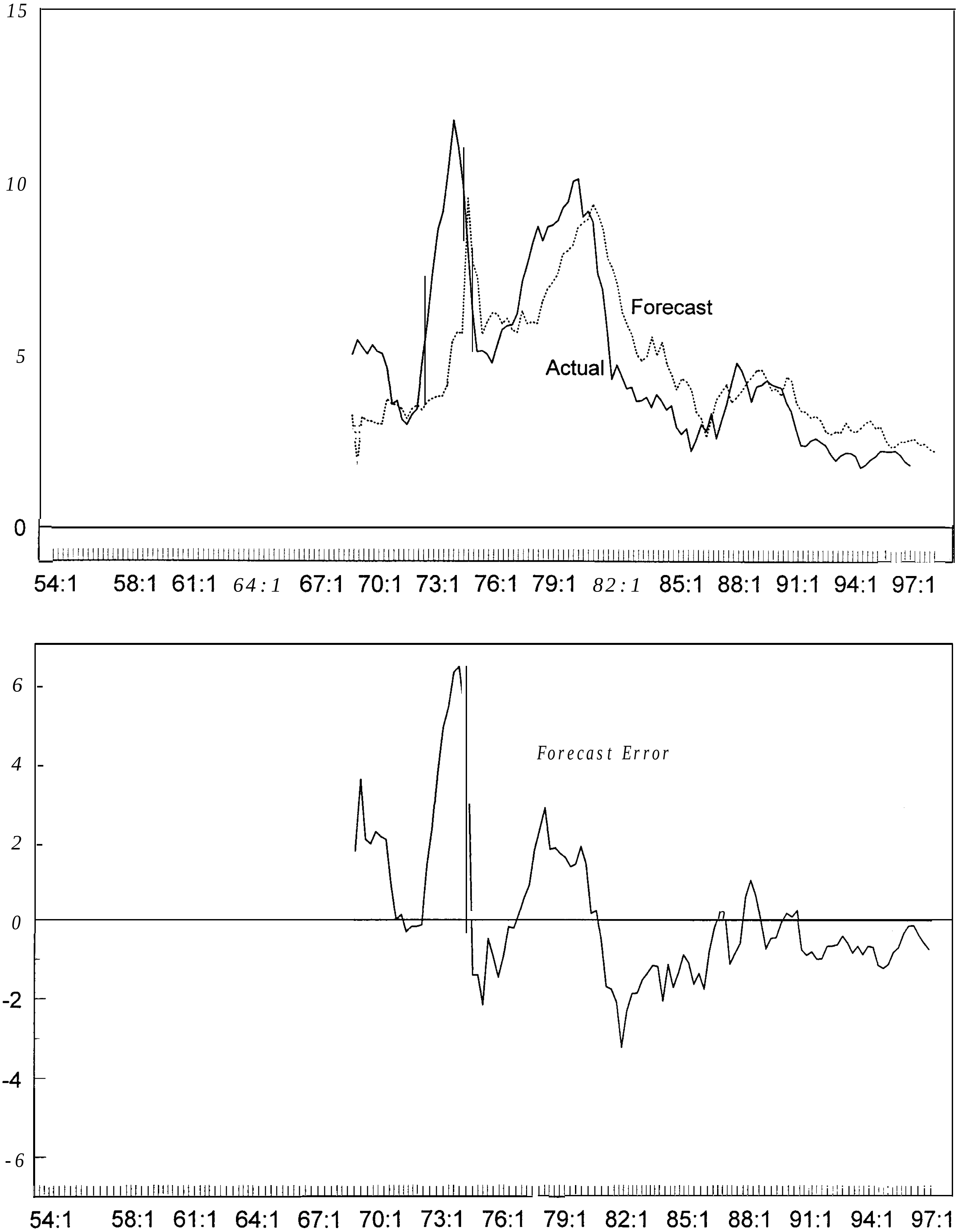
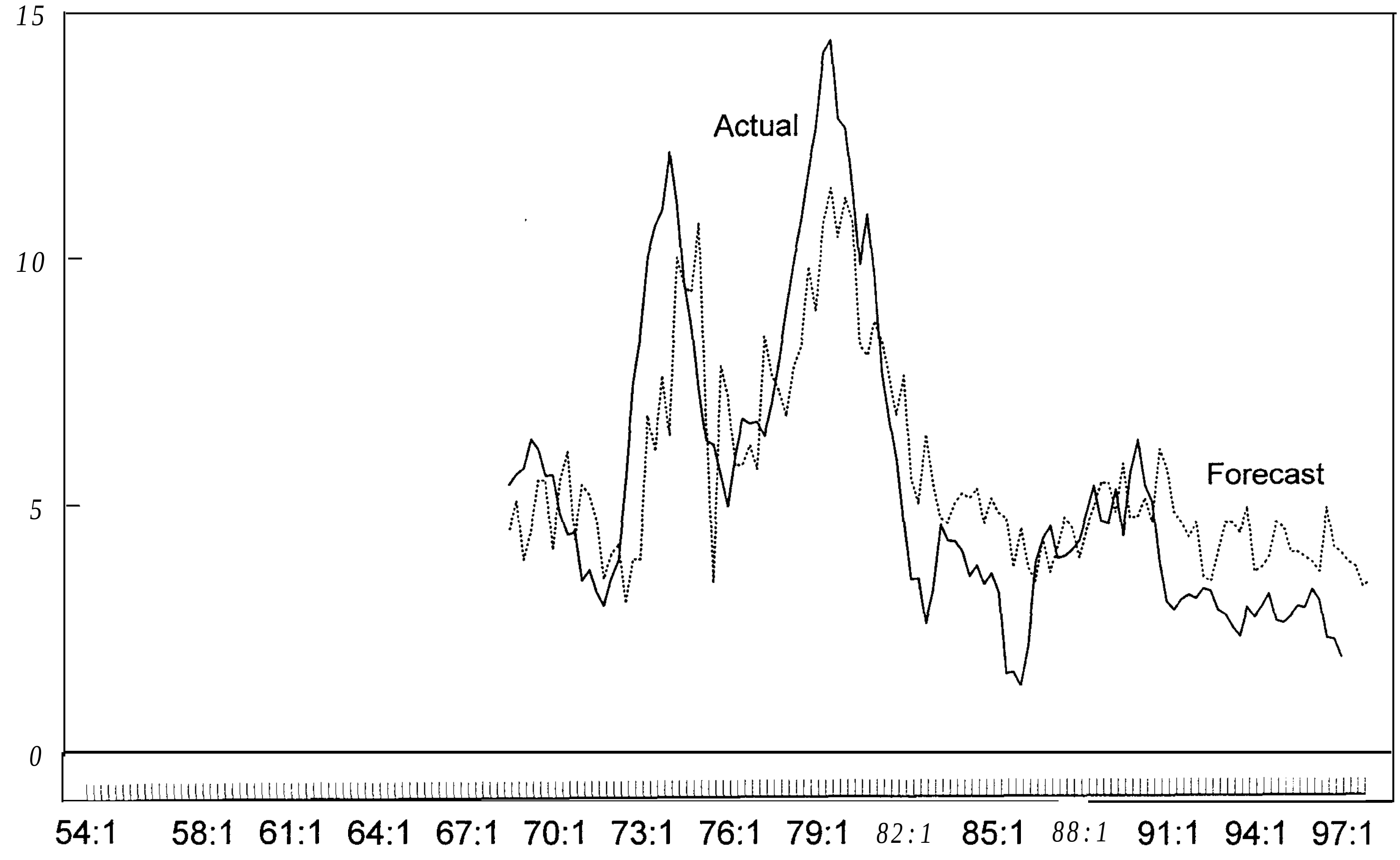


Figure 3
Michigan Survey



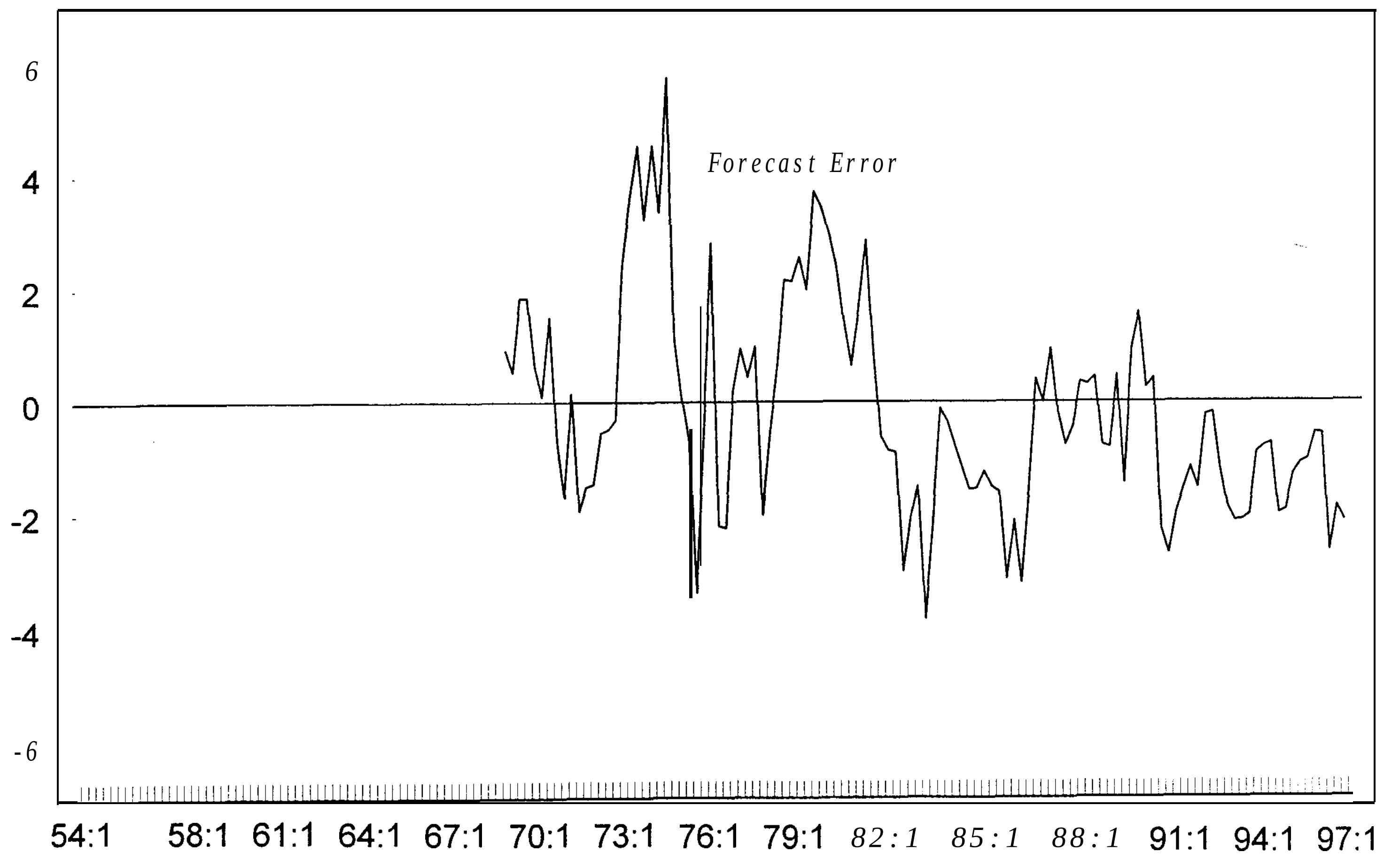


Figure 4
Livingston Forecast

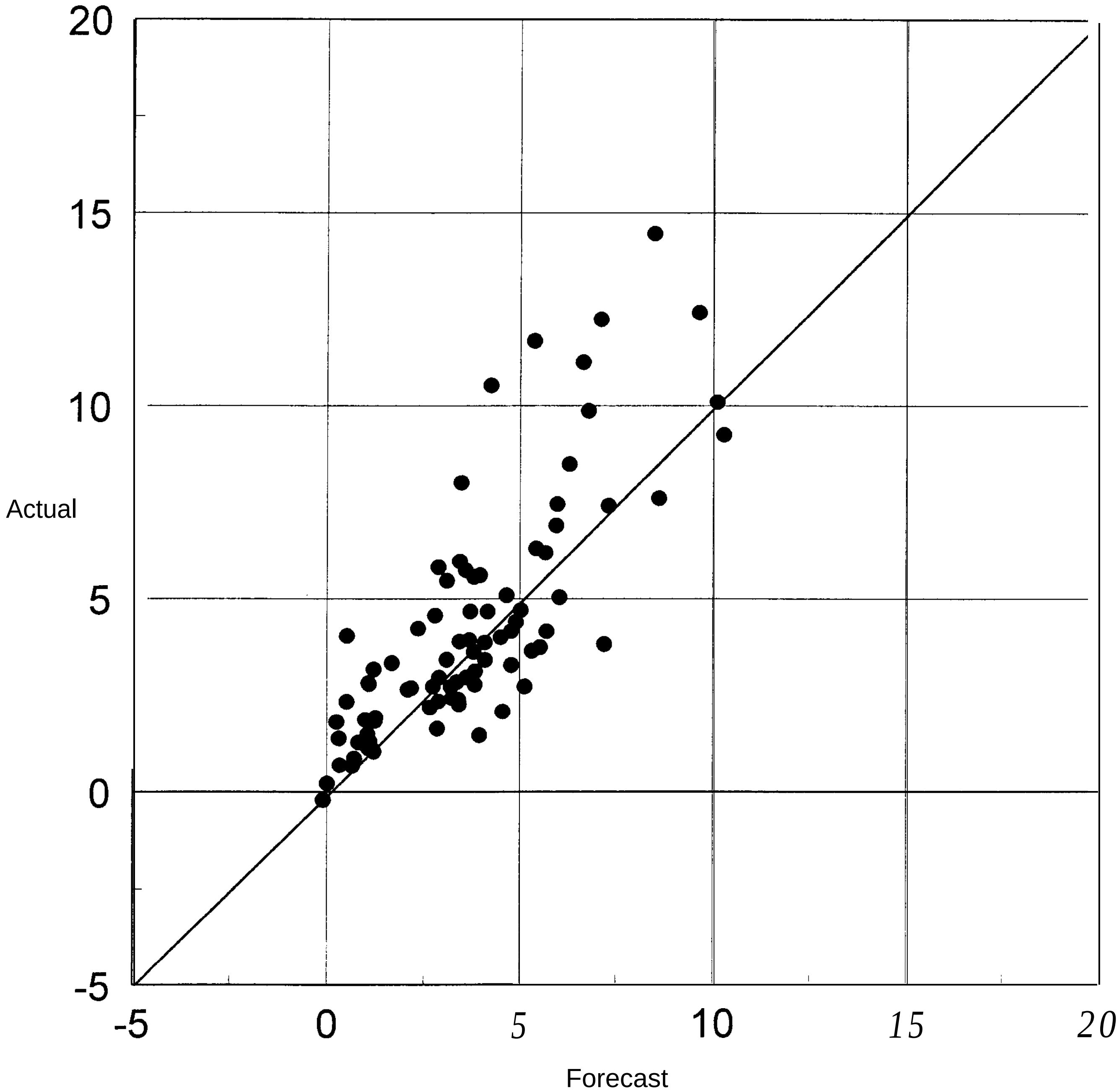


Figure 5
Survey of Professional Forecasters

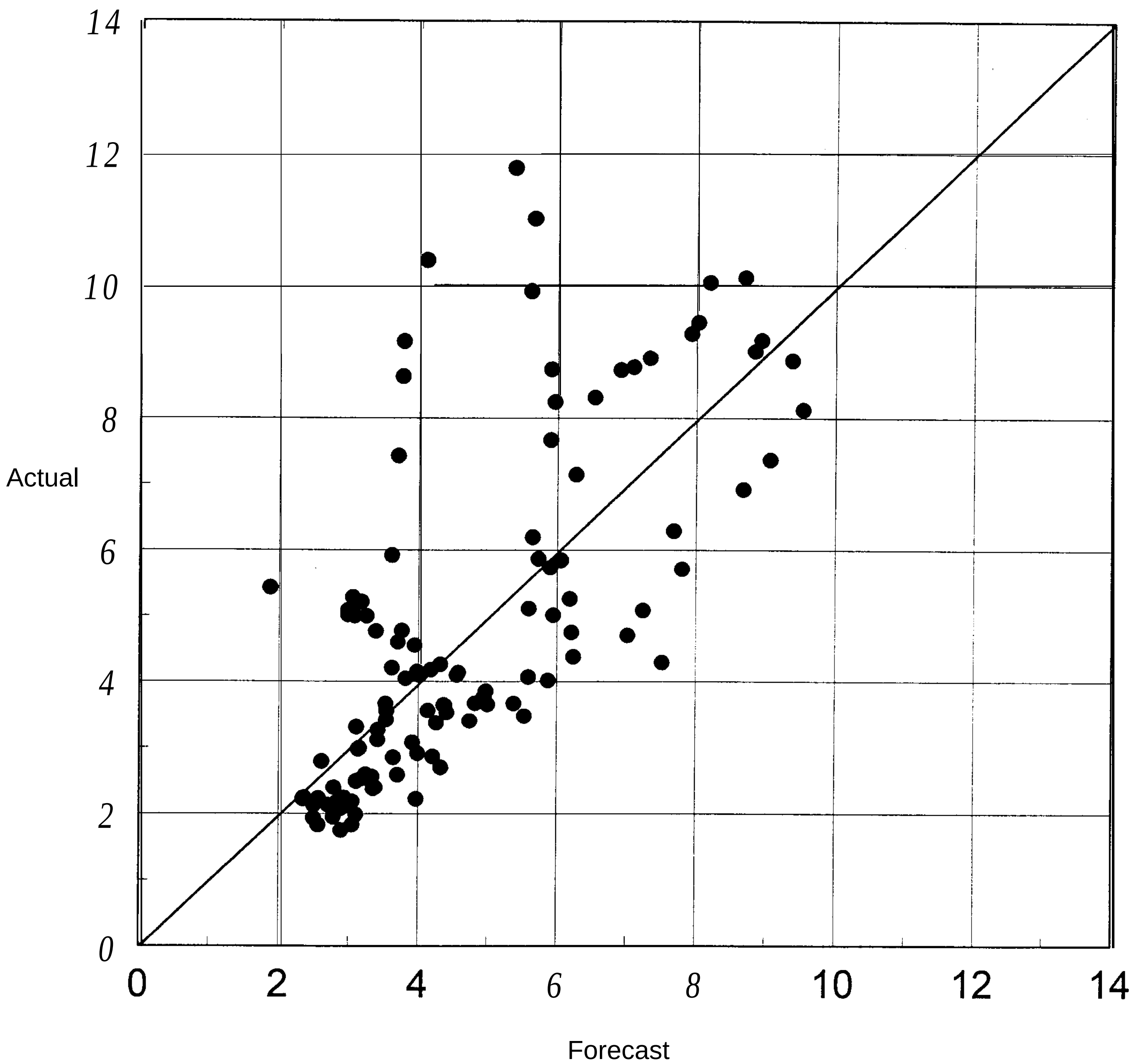


Figure 6
Michigan Survey