Research Program on Forecasting

Directional Forecasts of GDP and Inflation: A Joint Evaluation With an Application to Federal Reserve Predictions

Tara M. Sinclair
H.O. Stekler
Lindsay Kitzinger

RPF Working Paper No. 2006-002
http://www.gwu.edu/~forcpgm/2006-002.pdf

Updated Version
November 14, 2006

Note: A revised version of this paper has been accepted for publication in *Applied Economics.*

RESEARCH PROGRAM ON FORECASTING
Center of Economic Research
Department of Economics
The George Washington University
Washington, DC 20052
http://www.gwu.edu/~forcpgm

Directional Forecasts of GDP and Inflation: A Joint Evaluation With an Application to Federal Reserve Predictions[1]

Tara M. Sinclair
H.O. Stekler[2]
Lindsay Kitzinger

Department of Economics
George Washington University
1922 F Street NW
Old Main Suite 208
Washington DC 20052

November 14, 2006

JEL Codes: E37, C53, E52

Keywords: forecast evaluation, contingency tables, output forecast, inflation forecast

Abstract

Many studies have undertaken separate analyses of the Fed's forecasts of real GDP growth and inflation. This paper presents a method for jointly evaluating the direction of change predictions of these variables. We conclude that that some of the inflation forecasts, examined separately, were not valuable. However, the joint pattern of GDP and inflation projections was generally in accord with the economy's movements.

[1] Email: Sinclair tsinc@gwu.edu, Stekler hstekler@gwu.edu, Kitzinger lkitzinger37@gmail.com. The authors wish to thank Roy Bachelor, David Hendry, Fred Joutz, and participants in the 26th International Symposium on Forecasting for helpful comments and suggestions. All remaining errors are our own.
[2] Corresponding author. Email: hstekler@gwu.edu. Phone: 202-994-6150. Fax: 202-994-6147.

"…directional forecasting…is now an increasingly popular metric for forecasting performance…" (Pesaran and Timmermann, 2004, page 414).

1. Introduction

Private and public policymakers require accurate and valuable forecasts as inputs for their decisions. Many studies have evaluated economic predictions and provided information about the accuracy and value of these forecasts. While accuracy, as measured by quantitative errors, is important, it may be more crucial to accurately forecast the direction of change.[3] Will sales increase (decrease)? Will prices rise or fall? Will real GDP increase (decrease)? Will the rate of inflation increase (decrease)? In particular, the Federal Reserve monetary policy stance is often characterized as either expansionary (loose) or restrictive (tight). Given their importance, this paper will focus on direction of change forecasts. In particular we examine the situation when the decision-maker is required to forecast the direction of change of two related variables, such as sales and prices or, as in our case, real GDP and the inflation rate.

The procedures for evaluating directional forecasts often question whether these forecasts were "valuable" to decision makers. In this context, "value" is a test of the hypothesis that the forecast is better than a naïve no-change prediction. The procedure for testing whether forecasts have value for decision makers is based on the papers of Merton (1981) and Henriksson and Merton (1981). They focused on issues involving financial decision making. This methodology was introduced into the macroeconomic forecast evaluation literature by Schnader and Stekler (1990) and Stekler (1994) and has been used subsequently to test whether real GDP and inflation forecasts have value.

However, all of these studies have analyzed the forecasts of real GDP growth and inflation separately. Almost always forecasts for inflation and real GDP growth are made

[3] On this point see Leitch and Tanner (1991).

simultaneously by the same economists and are presented together. It is, therefore, appropriate and necessary to jointly evaluate the value of both forecasts made by the forecaster at the same time. This paper, therefore, extends the methodology that has been used to evaluate the directional accuracy of a forecast of a single variable and presents a method for jointly evaluating direction of change forecasts of two variables. It then applies this procedure to the GDP and inflation forecasts made by the staff of the Federal Reserve Board and published in the Greenbook.[4]

The next section explains the procedure for evaluating directional forecasts using 2x2 and 4x4 contingency tables. Section 3 will describe the data. The remaining sections present the results, the conclusions, and the implications of the results. Among the implications is the need to develop procedures for also evaluating *quantitative* macro forecasts beyond examining the GDP growth and inflation forecasts separately.

2. Evaluating Directional Forecasts

A. One Variable

Basically, when the real GDP and inflation forecasts are each evaluated separately, they are grouped into two categories. The GNP/GDP forecasts are categorized according to whether GDP growth was positive or negative, and the inflation categories depend on whether inflation increased or decreased.[5] A 2x2 contingency table is created that compares the predicted outcome of a variable with the actual outcome of that variable. (Table 1). For notation we have a total of N observations where for $n1$ of them both the actual and the predicted are positive and for $n2$ of them both the actual and the predicted are negative. We have n observations where the

[4] The Greenbook is the data and analytic summary notebook that is distributed before each meeting of the Federal Open Market Committee that determines monetary policy.

[5] No change is classified with the negative changes. Note that we are focusing on the direction of change in the inflation rate, which is equivalent to measuring accelerations and decelerations of the price level.

predicted outcome is positive and *N-n* observations where the predicted outcome is negative (or zero). We also have *N1* observations where the actual outcome is positive and $N2 = N - N1$ observations where the actual outcome is negative (or zero).

Table 1: The Relationship between Predicted and Actual Outcomes

Predicted Outcome	Actual Outcome		
	> 0	≤ 0	
> 0	n1	N2-n2	n
≤ 0	N1-n1	n2	N-n
	N1	N2	N

B. Joint Evaluation

The justification for jointly evaluating the Fed's GDP growth and inflation forecasts is that the Taylor (1993) rule contains these and only these variables for determining how much the Federal Reserve should change the interest rate.[6] As discussed by Woodford (2001), focusing only, but jointly, on inflation and output-gap stabilization as the goal of monetary policy has a sound theoretical basis. Moreover, forecasts of GDP growth and inflation are issued together and are related to each other. Often these forecasts come from the same forecasting model.

When the methodology is extended to examine the value of both forecasts simultaneously, a 2x2 contingency table cannot be used. Rather it will be necessary to use a 4x4 table.[7] In the expanded 4x4 table, instead of simply categorizing based on the separate GDP growth or inflation forecasts, the two types of forecasts are grouped into four categories: (1) GDP growth positive, inflation increasing, (2) GDP growth positive, inflation decreasing, (3) GDP growth

[6] Mishkin's popular money and banking textbook (Mishkin, 2006) says: "Caring about inflation and output fluctuations is consistent with many statements by Federal Reserve officials that c controlling inflation and stabilizing real output are important concerns of the Fed" (page 429).
[7] Naik and Leuthold (1986) also used a 4x4 contingency table in their qualitative analysis of forecasting performance. Their study focused on a different topic-the ability to predict turning points. (Also see Kaylen and Brandt, 1988).

negative, inflation increasing, and (4) GDP growth negative, inflation decreasing. The statistical tests are generalized versions of those used when the forecasts were analyzed separately. There is, however, a difference in interpretation once we go beyond the simple 2x2 case, as we discuss below.

C. Test Statistics

The statistical methodology tests whether or not the forecasts predict the associated directions of change. There are at least three test statistics that can be used to test the hypothesis that the forecasts fail to predict the observed events.[8] Two test statistics focus on independence. These test statistics are the Chi-square and Fisher's exact test. The Pesaran-Timmermann (1992) statistic specifically focuses on predictive failure. The forecasts are said to have value only if the null hypothesis of predictive failure is rejected. In the 2x2 case, the hypothesis of predictive failure is equivalent to the hypothesis that the actual and predicted values of the variable are not independent of each other. As discussed in Pesaran and Timmermann (1992), however, for the 4x4 case they are no longer equivalent. For our contingency table, independence implies predictive failure, but not vice versa.

The Chi-square test is the most common method used in evaluating contingency tables, but there are some problems with the assumptions required by the test. The first is that the Chi-square distribution is a continuous distribution while the test statistic is calculated using the discrete categories provided by the contingency tables. For 2x2 tables, there can be problems associated with simply using the Chi-square test since the approximation is not as accurate. The

[8] Merton (1981) and Henriksson and Merton (1981) had used a test based on the hypergeometric distribution. This is identical to Fisher's exact test. Their test assumes known row and column frequencies, which is not assumed for the Pesaran-Timmermann test.

solution for this problem is to use Yates' Continuity Correction. [9]

The test statistic is:

$$\chi^2 = \frac{\sum(|f_0 - f_e| - 0.5)^2}{f_e},$$

Where f_0 is the actual outcome and f_e is the expected outcome, measured as cell counts. It is important to note that Wickens (1989) warns that this test may be too conservative in the sense that the hypothesis of independence may be too often mistakenly accepted.

The other issue with the Chi-square test involves the expected frequency in a cell. One of the assumptions made under the Chi-square distribution's use as an approximation for the discrete frequencies of the contingency tables is that the expected frequencies in the cells are not too small. When the expected frequencies are small, this can cause the distribution to be an inaccurate approximation for the test statistic. For this reason, the general rule for the Chi-square test to be an accurate representation is that no more than 20% of the cells can have $f_e < 5$ and no cells should have $f_e < 1$ (Everitt, 1992). This posed a problem for many of the 4x4 tables.[10]

To avoid the problem of small expected frequencies associated with the 4x4 tables, Fisher's exact test was also calculated for each of the contingency tables. This method does not use the Chi-square distribution and does not calculate a test statistic. Rather, the hypergeometric distribution is used to directly calculate the probability of independence. The probability for a 2x2 table, given the notation of Table 1, is:

[9] With larger tables such as the 4x4 tables, the negative effect of applying the discrete categories to a continuous distribution is minimal because the number of categories (16) allows for a better approximation to the continuous distribution. It is, therefore, not necessary to use the correction and the standard Chi-square statistic can be used.
[10] See the tables in the Appendix for the 4x4 cell counts.

$$p = \frac{\binom{n}{n1}\binom{N-n}{N1-n1}}{\binom{N}{N1}} = \frac{N1!N2!n!(N-n)!}{n1!(N1-n1)!(N2-n2)!n2!N!}$$

We also calculated the Pesaran-Timmermann (1992) statistic. The general standardized test statistic for predictive performance for an *m* x *m* contingency table with a total of *N* observations is:

$$s_n = \sqrt{N}\mathbf{V}_n^{-1/2}S_n$$

where we have the following:

$$\hat{\mathbf{V}}_n = \left(\frac{\partial f(\mathbf{P})}{\partial \mathbf{P}}\right)_{\mathbf{P}=\hat{\mathbf{P}}} \hat{\mathbf{\Omega}} \left(\frac{\partial f(\mathbf{P})}{\partial \mathbf{P}}\right)_{\mathbf{P}=\hat{\mathbf{P}}}.$$

$\hat{P}_{ij} = n_{ij} / N$, where n_{ij} is the number of observations in the *ij* category of the contingency table.

$\hat{\mathbf{\Omega}} = \hat{\mathbf{\Psi}} - \hat{\mathbf{P}}\hat{\mathbf{P}}'$, where $\hat{\mathbf{\Psi}}$ is an m^2 x m^2 diagonal matrix with the elements of $\hat{\mathbf{P}}$ on the diagonal.

$S_n = \sum_{i=1}^{m}\left(\hat{P}_{ii} - \hat{P}_{i0}\hat{P}_{0i}\right)$, where $\hat{P}_{i0} = n_{i0} / N$.

$\hat{P}_{0i} = n_{0i} / N$, where n_{i0} and n_{0i} represent the i^{th} row and column totals respectively.

Pesaran and Timmermann (1992) present their results based on the square of this test statistic in order to more easily compare it to the Chi-square goodness of fit statistic. We will thus report s_n^2 as the Pesaran-Timmermann test statistic in the tables below. This test statistic has a Chi-square distribution with one degree of freedom.

Pesaran and Timmermann's predictive-failure test is particularly useful in the case where we undertake a joint evaluation of GDP growth and inflation forecasts. Their test does not require that the two forecasts be independent of each other. Since output and inflation may be predicted from the same forecasting model, this is an important consideration.

3. The Data

At least once every quarter, the Federal Open Market Committee (FOMC) meets to discuss and evaluate the economic situation of the United States and to set monetary policy. At these meetings, the Federal Reserve's Greenbook is distributed and used as the basis for discussion. Since 1965 the Greenbook has contained quantitative forecasts of inflation, real GDP growth, and other economic factors. These forecasts are released to the public after five years.

The data set consists of 1262 observations of both real GDP growth and inflation (measured as the percentage change of the GDP deflator) spanning a time period from 1966.1 to 1997.4. While the Greenbook contains forecasts made from 0-8 quarters ahead, this paper will only examine forecasts made for the current quarter and one quarter ahead. We focus on short horizons because the forecasts in the Greenbook sometimes assume that the interest rate path would remain constant. Since changes in monetary policy are not likely to affect economic activity less than two quarters ahead, this assumption does not bias those forecasts. The assumption may produce biases in the longer term forecasts.

There are multiple observations per quarter. If the forecast for the current quarter were made in the last month of the quarter, it was considered to have a lead of zero months. At the same time that this forecast was issued for the current quarter, another one was made for the one quarter-ahead period. The forecast horizon in this case is three months. Thus current quarter forecasts had leads of zero, one and two months, while the one quarter-ahead forecasts had leads of three, four and five months.

It is important to note that the FOMC met more frequently per quarter in the 1960s and 1970s. As a result, there are often multiple forecasts per lead time per quarter for many of the years spanning the 1969-1980 period. There are fewer forecasts from the 1980s and 1990s.

We assume that the objective of the forecaster is to forecast the data released by the Bureau of Economic Analysis (BEA) approximately 45 to 60 days after the end of the quarter.[11] We, therefore, use these data as the "actual" data with which the Greenbook forecasts are compared. A theoretical justification for using these data was provided by Zarnowitz (1985). He argued that the revised data issued later may not be "conceptually comparable" to the Greenbook for two reasons. They not only incorporate new information, but the BEA also makes definitional and classification changes that the forecaster presumably could not foresee.

4. Results

A. Separate Evaluations

We first examined the GDP growth and inflation forecasts separately to determine whether each one was valuable. This enabled us to compare our results with those of Joutz and Stekler (2000) using our larger data set. Table 2 presents the 2x2 contingency tables for the separate GDP growth and inflation forecasts. We then calculated the three statistics to test the null that the forecasts and outcomes were independent.[12] The associated probabilities are shown in Table 3.

---

[11] Our sample spans several terminology changes at the BEA. For the beginning of our sample the "final" data were released 45 days after the end of the quarter with no further revisions until more comprehensive revisions were made that might include definitional or classificational changes. Starting in 1974, these 45-day numbers were also referred to as 1st revision numbers because a second revision was introduced approximately 75 days after the end of the quarter. Since 1988 the data we use have been referred to by the BEA as "preliminary" data and are released approximately two months after the end of the quarter (with "final" data being released approximately 3 months after the end of the quarter). The only real-time data consistently available throughout our sample are the 45 to 60 day numbers, so we focus on these for analysis.

[12] Note that in the 2x2 case, the Pesaran test statistic is equivalently a test of independence and of predictive ability.

Table 2: Cell Counts for 2x2 Contingency Tables										
	Real GDP growth					**Δ Inflation**				
		P > 0	P ≤ 0	P > 0	P ≤ 0		P > 0	P ≤ 0	P > 0	P ≤ 0
Lead (months)	N	A > 0	A ≤ 0	A ≤ 0	A > 0	N	A > 0	A ≤ 0	A ≤ 0	A > 0
0	139	113	15	6	5	139	55	53	19	12
1	97	76	10	5	6	97	34	36	14	13
2	85	66	10	6	3	85	30	24	16	15
3	140	110	11	11	8	140	43	46	26	25
4	92	71	5	9	7	92	24	30	20	18
5	78	61	6	6	5	78	23	23	12	20

Table 3: Probability of Null Hypothesis, GDP growth and Inflation Separately						
	Real GDP growth			**Δ Inflation**		
Lead (months)	Chi-Square (Yates Correction)	Fisher Exact	Pesaran-Timmermann	Chi-Square (Yates Correction)	Fisher Exact	Pesaran-Timmermann
0	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
1	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
2	< 0.001	< 0.001	< 0.001	0.025	0.017	0.011
3	< 0.001	< 0.001	< 0.001	0.002	0.002	0.001
4	0.021	0.017	0.061	0.153	0.142	0.097
5	< 0.001	0.001	0.015	0.142	0.112	0.083

First looking at real GDP growth, we find only one exception to the results found by Joutz and Stekler (2000). They had found that, at all lead times, the GDP growth forecasts were valuable both for the current and future quarter. The one exception is that in the case of the 4 month lead the Pesaran-Timmermann statistic rejects predictive failure beyond the .06 level.

The results for inflation were similar but not identical to the results of Joutz and Stekler (2000). They had found that current quarter inflation forecasts made at all horizons were valuable; however, for all leads of the one quarter ahead inflation predictions they were not able to reject independence. According to our analysis, all the inflation forecasts for the current quarter as well as the one with a lead of three months were found to be valuable by all three tests. At the .05 level, all the test statistics fail to reject the null that the four and five month directional inflation forecasts are independent of the observed increases and decreases in the rate of inflation. These results stand in stark contrast to the findings of Romer and Romer (2000) and Sims (2002). They had found that the quantitative forecasts of the Fed were rational and at least as good as those of private forecasters. Our interpretation of the failure to associate the predictions of increases and decreases in the inflation rate with the actual changes in that rate means that these inflation forecasts, by themselves, were not valuable to the decision makers at the Federal Reserve in their conduct of monetary policy.

B. Joint Evaluation

We now turn to the joint evaluation of the GDP and inflation forecasts. The probabilities derived from the test statistics are presented in Table 4. The 4x4 contingency tables are presented in the Appendix. The results indicate that, with one exception, the combined forecasts for both the current quarter and future quarter (for all lead times from zero to five months) can be said to be valuable. The exception comes from the Pesaran-Timmermann statistic which is not significant at the .05 level for the four month lead.[13]

[13] One assumption of the Pesaran-Timmermann test statistic is that the probability of changes in direction must be time-invariant. Due to the longer expansions since the mid-1980s, this may not be the case, in particular for GDP growth. To determine the effect of the longer GDP expansions, we split the sample pre- and post-1984. The results were similar to the full sample except for leads 4 and 5. For lead 4 there were only 29 observations (and only one recession) in the post-1984 subset which may explain the lack of significance for the Pesaran-Timmermann test statistic for lead 4 for this subset. For lead 5, the test statistic is not significant at the 5% level for either subset, even though for the full sample it is significant. This occurs because the off-diagonal elements are clustered differently in the two subsamples. These results are available from the authors upon request.

Table 4: Probabilities Associated with 3 Test Statistics for 4x4 Contingency Table			
Lead (months)	Chi-Square	Fisher Exact	Pesaran-Timmermann
0	< 0.001	< 0.001	< 0.001
1	< 0.001	< 0.001	< 0.001
2	< 0.001	< 0.001	< 0.001
3	< 0.001	< 0.001	< 0.001
4	0.01	0.01	0.08
5	0.001	< 0.001	0.02

Despite the failure to always associate the increases and decreases of the inflation rate with the observed changes, the joint pattern of GDP and inflation projections was generally in accord with the economy's actual performance. Although the inflation forecasts, *by themselves*, were found to not always be valuable, the *joint* forecasts appear to have been valuable overall.

5. Conclusions and Implications

This paper has shown that the Fed's inflation forecasts by themselves might not have been useful to the Federal Reserve decision makers. However, we also developed a method for undertaking a joint qualitative evaluation of GDP growth and inflation. It is no longer necessary to analyze the directional characteristics of the GDP growth and inflation forecasts separately.

When this method was applied to the Greenbook forecasts of the Federal Reserve, we found that on average the forecasts for the current quarter and one quarter-ahead period yielded an accurate view of the state of the economy. However, there is a caveat. The method gives equal weight to forecasts made in periods when forecasting was easy and to periods where prediction of the economy was more difficult. The forecasts of most individuals and groups have been less accurate in the vicinity of cyclical turns. If there are long periods of positive periods of economic GDP growth with stable inflation that are easy to predict, this method would make it more likely that the null of independence or predictive failure would be rejected. This was the case in the 1990s and may explain why our results, analyzing the inflation forecasts alone, differ slightly from those of Joutz and Stekler. They had found that all future quarter inflation forecasts were not shown to be valuable, while our results show that this finding applies only to forecasts with a lead of four months. Their data set ended in mid 1989, while ours ends at the end of 1997.

Finally, we must note that this procedure for evaluating GDP growth and inflation forecasts jointly only applies to qualitative directional predictions. This evaluation method only considers a limited amount of information, i.e. the direction of change. The procedure does not take into account the magnitude of any error associated with individual forecasts made by the Federal Reserve. A procedure for evaluating quantitative predictions jointly remains to be developed.

References

Everitt, B.S., (1992). *The Analysis of Contingency Tables, Second Edition*. London: Chapman and Hall.

Henriksson, R.D. and Merton R.C., (1981), On market timing and investment performance II: Statistical procedures for evaluating forecasting skills, *Journal of Business* 54, 513-523.

Joutz, F. and Stekler H.O., (2000), An evaluation of the predictions of the Federal Reserve, *International Journal of Forecasting,* 16, 17-38.

Leitch, G and Tanner, E.J., (1991), Economic forecast evaluations: Profits versus conventional error measures, *American Economic Review,* 81, 580-590.

Kaylen, M.S. and Brandt, J. A., (1988), A note on qualitative forecast evaluation: Comment, *American Journal of Agricultural Economics,* 70, 415-16.

Merton,R.C., (1981), On market timing and investment performance I: An equilibrium theory of value for market forecasts, *Journal of Business* 54, 363-406.

Mishkin, F.S., (2006), *The Economics of Money Banking and Financial Markets*, New York, Pearson.

Naik, G. and Leuthold, R. M., (1986), A note on qualitative forecast evaluation, *American Journal of Agricultural Economics,* 68, 721-26.

Pesaran, M. H. and Timmermann, A., (1992), A simple nonparametric test of predictive performance, *Journal of Business & Economic Statistics,* 10, 461-465.

Pesaran, M. H. and Timmermann, A., (2004), How costly is it to ignore breaks when forecasting the direction of a time series? *International Journal of Forecasting,* 20, 411-425.

Romer, C. D. and Romer, D. H., (2000), Federal Reserve information and the behavior of interest rates, *American Economic Review,* 90, 429-457.

Schnader, M.H. and Stekler, H.O., (1990), Evaluating predictions of change, *Journal of Business,* 63, 99-107.

Sims, C. A., (2002), The role of models and probabilities in the monetary policy process. *Brookings Papers on Economic Activity* 2, 1-40.

Stekler, H.O., (1994), Are economic forecasts valuable? *Journal of Forecasting* 13, 495-505.

Taylor, J.B., (1993), Discretion versus policy rules in practice, *Carnegie-Rochester Conference Series on Public Policy,* 39,195-214.

Wickens, T. D., (1989), *Multiway Contingency Tables Analysis for the Social Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Woodford, M., (2001), The Taylor rule and optimal monetary policy, *American Economic Review*, 91 (2), 232-237.

APPENDIX

Table A1: The 4x4 Contingency Table for the Zero Month Lead

	Actual Outcome			
	GDP > 0, Δinf > 0	GDP > 0, Δinf ≤ 0	GDP ≤ 0, Δinf > 0	GDP ≤ 0, Δinf ≤ 0
Predicted Outcome				
GDP > 0, Δinf > 0	49	13	1	1
GDP > 0, Δinf ≤ 0	7	43	0	4
GDP ≤ 0, Δinf > 0	1	2	4	2
GDP ≤ 0, Δinf ≤ 0	0	3	5	4

Table A2: The 4x4 Contingency Table for the One Month Lead

	Actual Outcome			
	GDP > 0, Δinf > 0	GDP > 0, Δinf ≤ 0	GDP ≤ 0, Δinf > 0	GDP ≤ 0, Δinf ≤ 0
Predicted Outcome				
GDP > 0, Δinf > 0	28	11	2	1
GDP > 0, Δinf ≤ 0	9	28	0	2
GDP ≤ 0, Δinf > 0	2	0	2	2
GDP ≤ 0, Δinf ≤ 0	0	4	4	2

Table A3: The 4x4 Contingency Table for the Two Month Lead

	Actual Outcome			
	GDP > 0, Δinf > 0	GDP > 0, Δinf ≤ 0	GDP ≤ 0, Δinf > 0	GDP ≤ 0, Δinf ≤ 0
Predicted Outcome				
GDP > 0, Δinf > 0	27	15	1	0
GDP > 0, Δinf ≤ 0	8	16	3	2
GDP ≤ 0, Δinf > 0	0	1	2	0
GDP ≤ 0, Δinf ≤ 0	1	1	3	5

Table A4: The 4x4 Contingency Table for the Three Month Lead				
	Actual Outcome			
	GDP > 0, Δinf > 0	GDP > 0, Δinf ≤ 0	GDP ≤ 0, Δinf > 0	GDP ≤ 0, Δinf ≤ 0
Predicted Outcome				
GDP > 0, Δinf > 0	35	23	1	2
GDP > 0, Δinf ≤ 0	19	33	3	5
GDP ≤ 0, Δinf > 0	4	1	3	0
GDP ≤ 0, Δinf ≤ 0	0	3	3	5

Table A5: The 4x4 Contingency Table for the Four Month Lead				
	Actual Outcome			
	GDP > 0, Δinf > 0	GDP > 0, Δinf ≤ 0	GDP ≤ 0, Δinf > 0	GDP ≤ 0, Δinf ≤ 0
Predicted Outcome				
GDP > 0, Δinf > 0	20	18	1	2
GDP > 0, Δinf ≤ 0	13	20	3	3
GDP ≤ 0, Δinf > 0	3	0	0	0
GDP ≤ 0, Δinf ≤ 0	0	4	2	3

Table A6: The 4x4 Contingency Table for the Five Month Lead				
	Actual Outcome			
	GDP > 0, Δinf > 0	GDP > 0, Δinf ≤ 0	GDP ≤ 0, Δinf > 0	GDP ≤ 0, Δinf ≤ 0
Predicted Outcome				
GDP > 0, Δinf > 0	21	12	1	0
GDP > 0, Δinf ≤ 0	13	15	4	1
GDP ≤ 0, Δinf > 0	0	0	1	0
GDP ≤ 0, Δinf ≤ 0	0	5	3	2