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

Tara M. Sinclair, H.O. Stekler, and Lindsay Kitzinger

Department of Economics
The George Washington University

Brown Bag Seminar Series on Forecasting
January 9, 2007

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON DC

Motivation

"…directional forecasting…is now an increasingly popular metric for forecasting performance…"
--Pesaran and Timmermann, IJF 2004.

- Directional forecasts matter for both private and public policymakers.
- In particular, the Federal Reserve monetary policy stance is often characterized as either expansionary (loose) or restrictive (tight).

Motivation 2

- Almost always forecasts for inflation and real GDP growth are made simultaneously by the same economists and are presented together.
- Previous studies, however, have analyzed the directional forecasts of real GDP growth and inflation separately.
- We instead propose to evaluate them jointly.

Outline of the Talk

- Methodology for Evaluating Directional Forecasts
  - The 2x2 contingency table
  - Joint evaluation: the 4x4 contingency table
  - Test Statistics
- Application:
  Are the Fed's Forecasts Jointly Valuable?
  - Data
  - Results
- Conclusions and Implications
- Extensions

Evaluating Directional Forecasts

- We define forecasts as “valuable” if they perform better than the naïve no-change prediction.
  - For joint evaluation, we focus on rejecting predictive failure.
- For our application, we will evaluate the performance of directional forecasts of the change in real GDP and the change in inflation.
  - whether real GDP growth (the change in GDP) was positive or negative.
  - Whether inflation increased or decreased (whether the change in inflation was positive or negative).
  - Examining the direction of change provides sufficient positive and negative observations for analysis.

The 2x2 Contingency Table

- Consider evaluating GDP growth by itself.
- GDP growth can be either positive or negative (group no-change with negative).
- The forecaster has two possible forecasts: positive or negative.
- The actual outcome has two possibilities: positive or negative.
- This leads to a 2x2 contingency table.

The 2x2 Contingency Table

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

N: Total Observations
n: Total Predicted Positive
N1: Total Actual Positive
N2: Total Actual Negative (or zero)
n1: Total Positive for both Predicted and Actual
n2: Total Negative (or zero) for both Predicted and Actual

Example: Real GDP Growth

Table 2a: The 2x2 Contingency Table for Real GDP Growth for the Zero Month Lead			
	Actual Outcome		
Predicted Outcome	Real GDP Growth > 0	Real GDP Growth ⩽ 0	
Real GDP Growth > 0	113	6	119
Real GDP Growth ⩽ 0	5	15	20
	118	21	139

The 4x4 Contingency Table

- Now consider jointly evaluating forecasts of GDP growth and the change in inflation.
- The forecaster and the actuals now each have four possibilities:
  1) GDP growth positive, inflation increasing
  2) GDP growth positive, inflation decreasing
  3) GDP growth negative, inflation increasing
  4) GDP growth negative, inflation decreasing
- This leads to a 4x4 contingency table.
- The 4x4 contingency table has not previously been used in the literature for forecast evaluation.

The 4x4 Contingency Table

Table 1a: The Relationship between Predicted and Actual Outcomes					
Predicted Outcome	Actual Outcome				
	GDP > 0, Δinf > 0	GDP > 0, Δinf ≤ 0	GDP ≤ 0, Δinf > 0	GDP ≤ 0, Δinf ≤ 0	
GDP > 0, Δinf > 0	n1	n1,2	n1,3	n1,4	n1,0
GDP > 0, Δinf ≤ 0	n2,1	n2	n2,3	n2,4	n2,0
GDP ≤ 0, Δinf > 0	n3,1	n3,2	n3	n3,4	n3,0
GDP ≤ 0, Δinf ≤ 0	n4,1	n4,2	n4,3	n4	n4,0
	N1	N2	N3	N4	N

N: Total Observations
N1 thru N4: Column Totals
n1,0 thru n4,0: Row Totals
n1 thru n4: Predicted matches Actual

Example: 4x4 Contingency Table

Table A1: The 4x4 Contingency Table for the Zero Month Lead				
	Actual Outcome			
	Δ GDP > 0, Δ inf > 0	Δ GDP > 0, Δ inf $\leqslant$ 0	Δ GDP $\leqslant$ 0, Δ inf > 0	Δ GDP $\leqslant$ 0, Δ inf $\leqslant$ 0
Predicted Outcome				
Δ GDP > 0, Δ inf > 0	49	13	1	1
Δ GDP > 0, Δ inf $\leqslant$ 0	7	43	0	4
Δ GDP $\leqslant$ 0, Δ inf > 0	1	2	4	2
Δ GDP $\leqslant$ 0, Δ inf $\leqslant$ 0	0	3	5	4

Test Statistics

- The statistical methodology tests whether or not the forecasts predict the associated directions of change.
- For the 2x2 case, the hypothesis of predictive failure is equivalent to the hypothesis of independence.
- For the 4x4 case, independence implies predictive failure, but not vice-versa.

Three Test Statistics

- Two test statistics focus on independence:
  - Chi-square test.
  - Fisher's exact test.
- The third test statistic focuses on predictive failure:
  - Pesaran and Timmermann (1992)

Chi-Square Test

- The Chi-square test is the most common method used in evaluating contingency tables.
- Drawbacks:
  - Chi-square distribution is a continuous distribution while the test statistic is calculated using discrete categories.
    - Use the Yates' Continuity Correction for 2x2.
  - The test may be too conservative in the sense that independence may not be rejected often enough (Wickens, 1989).
  - Requires expected frequencies in the cells to not be too small for standard distribution of the test statistic (a problem for the off-diagonals, particularly in the 4x4 case).

Fisher's Exact Test

- Fisher's Exact Test avoids the problem of small expected frequencies.
- This method uses the hypergeometric distribution to directly calculate the probability of independence.

Pesaran and Timmermann's Test

- Pesaran and Timmermann (1992) propose a more appropriate test statistic for our joint forecast evaluation.
  - Tests predictive failure instead of independence.
  - Does not require that the two forecasts be independent of each other.

Application: Are the Fed's Forecasts Jointly Valuable?

- Evaluating the Fed's directional forecasts of GDP growth and inflation changes.
- Joint evaluation: the two forecasts often come from the same forecasting model.
- Only inflation and GDP: they are the only two included in the Taylor Rule.

Forecast Data

- Greenbook forecasts of inflation (based on GDP deflator) and real GDP growth
- 1262 observations from the first quarter of 1966 through the 4th quarter of 1997.
  - Multiple observations per quarter depending on the number of FOMC meetings that quarter.
    - The FOMC met more frequently per quarter in the 1960s and 1970s than later in the sample.
- We only examine forecasts for the current quarter and 1 quarter ahead.
  - Focus on short horizons to avoid the effect of any changes in monetary policy.

Leads

Forecast Date	Current Quarter Forecast Lead	One-Quarter-Ahead Forecast Lead
First month of quarter	Two month lead	Five month lead
Second month of quarter	One month lead	Four month lead
Third month of quarter	Zero month lead	Three month lead

Actual Outcome Data

- Assume the objective is to forecast data released 45-60 days after the end of the quarter.
  - Avoids definitional and classification changes.
  - Terminology for these data releases varied over the sample:
    - Before 1974, the “final” data: 45 days after the end of the quarter.
    - Starting in 1974, “1st revision” (second revision about 75 days out).
    - Since 1988, the “preliminary” data are released approximately two months after the quarter.

2x2 Results

Table 3: Probability of Null Hypothesis, GDP Growth and ΔInflation Separately						
	Real GDP growth			**Δ Inflation**		
Lead	Yates Chi-Square	Fisher Exact	P-T	Yates Chi-Square	Fisher Exact	P-T
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

Comparison with Joutz-Stekler (2000)

- Real GDP Growth
  - Joutz and Stekler found forecasts were valuable at all six lead times.
  - We found all except one: the Pesaran Timmermann statistic did not reject for lead 4.
- Inflation Changes
  - Joutz and Stekler found that only current quarter forecasts were valuable (leads 0 thru 2).
  - We found that lead 3 was also valuable, but not 4 or 5.

4x4 Results

Table 4: Probabilities for 4x4 Contingency Table			
Lead	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

Interpreting 4x4 Results

- Only one exception where the forecasts were not jointly valuable.
- Inflation forecasts *by themselves* are not always valuable (particularly at longer leads).
- But, the joint pattern of GDP and inflation direction of change forecasts was generally in accord with the economy's actual performance.

Conclusions and Implications

- We developed a simple method for joint evaluations of directional forecasts.
- It appears that forecasts by the Fed of GDP and inflation are in general informative about the true state of the economy.
- A caveat: The method gives equal weight to forecasts made at any point in time.
  - Forecasts may be more difficult around turning points.

Extensions

- New work underway by Sinclair, Stekler, and Reid: A procedure for jointly evaluating **quantitative** predictions.
- We also need procedures for testing for joint rationality.
  - Hanson and Whitehorn (2006)
  - Work underway by Sinclair and Stekler.
  - Also work underway by Ivana Komunjer (UCSD) and Michael Owyang (STL Fed).