Journal of Economics and Business xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Journal of Economics and Business

Comparing Federal Reserve, Blue Chip, and time series forecasts of US output growth

Hamid Baghestani, Bassam M. AbuAl-Foul*

Department of Economics, School of Business Administration, American University of Sharjah, P.O. Box 26666, Sharjah, United Arab Emirates

ARTICLE INFO

Article history:
Received 22 January 2016
Received in revised form 3 November 2016
Accepted 10 November 2016
Available online xxx

JEL classification:
E22
E32
R31
R32

Keywords:
Asymmetric information
Residential investment
Consumer sentiment
Forecast evaluation

ABSTRACT

We evaluate the predictive content of Federal Reserve and Blue Chip forecasts of output growth by utilizing two comparable forecasts as benchmarks: a univariate autoregressive (AR) model, and a vector autoregressive (VAR) model which includes output growth, growth in residential investment, and consumers' assessments of business conditions. We first show the forecasts are all directionally accurate, free of systematic bias, and efficient. Second, the asymmetric information hypothesis cannot be supported. Third, the Federal Reserve and private forecasts are generally less informative than the VAR forecasts and thus lack past information on residential investment growth and consumers' assessments of business conditions.

1. Introduction

Despite the inherent difficulty, both public and private forecasters are regularly engaged in predicting output growth. Market participants seek accurate forecasts of growth for making a variety of economic and financial decisions including investment. Such forecasts are also key inputs for both fiscal and monetary authorities in formulating economic policies (Chauvet and Potter, 2013). In evaluating the accuracy of output growth, inflation, and unemployment forecasts, studies have often tested the asymmetric information hypothesis that the Federal Reserve has useful information about the state of the economy that is not known by the private sector. Romer and Romer (2000), Gavin and Mandal (2001), and Sims (2002) convincingly support this hypothesis for inflation forecasts. However, as noted by Gavin and Mandal (2001), the findings are rather weak for output growth forecasts. In addition, Baghestani (2008) shows that the private forecasts of unemployment are more informative than the Federal Reserve forecasts.[1]

In this study, we evaluate the predictive information content of the Federal Reserve and private (Blue Chip) forecasts of output growth by employing two sets of comparable forecasts as benchmarks. The first set is from a univariate autoregressive

* Corresponding author.
E-mail addresses: baghesta@msn.com (H. Baghestani), babufoul@aus.edu (B.M. AbuAl-Foul).

[1] Baghestani (2011) investigates the predictive information content of the Federal Reserve and private forecasts of non-residential and residential investment and reports mixed evidence in support of the asymmetric information hypothesis.

http://dx.doi.org/10.1016/j.jeconbus.2016.11.001

(AR) model, and the second one is from a vector autoregressive (VAR) model. The AR forecasts contain past information in output growth, and the VAR forecasts contain past information on output growth, growth in residential investment, and consumers' assessments of business conditions.

There are two noteworthy aspects to this study. First, we utilize real time data to provide out-of-sample evidence on the usefulness of growth in residential investment for predicting output growth. This complements existing studies which have provided in-sample evidence. For instance, Green (1997) utilizes the Granger-causality approach to demonstrate that, unlike non-residential investment, residential investment Granger-causes GDP. As demonstrated by Coulson and Kim (2000), the reason behind such evidence is that, unlike non-residential investment, residential investment has a significant impact on consumption. Leamer (2007) shows that "It is residential investment that contributes most to weakness before recessions." Toward a more effective monetary policy, Leamer argues for a new Taylor Rule in which GDP is replaced by housing leading indicators.[2]

Second, the inclusion of consumers' assessments of business conditions in the VAR model is important in light of existing literature that offers mixed results. Carroll, Fuhrer, and Wilcox (1994) and Bram and Ludvigson (1998) present evidence in support of consumer sentiment as a reliable predictor of consumption growth. Croushore (2005) replicates these studies using real-time data and finds that consumer sentiment is of little value. Garner (1991) argues that consumer sentiment is rarely a useful predictor of economic performance. Batchelor and Dua (1998) show that consumer sentiment could predict only the 1991 US recession. More recent studies by Dees and Brinca (2013), Christiansen, Eriksen, and Møller (2014), and Österholma (2014), however, find that consumer sentiment has significant predictive power for economic indicators. The study by Christiansen et al. (2014), in particular, shows that sentiment indexes can significantly help improve predictions of US recessions.

We find a number of important results. First, the Federal Reserve, Blue Chip, AR, and VAR forecasts are all directionally accurate, free of systematic bias, and efficient. Second, our test results do not support the asymmetric information hypothesis that the Federal Reserve has useful information about the state of the economy that is not known by the private sector. Third, the VAR forecasts embody useful predictive information above and beyond that contained in the AR forecasts. This means that past information on growth in residential investment and consumers' assessments of business conditions is useful in predicting output growth. Fourth, the VAR forecasts generally embody useful predictive information beyond that contained in the Federal Reserve and Blue Chip forecasts. This suggests that the Federal Reserve and Blue Chip forecasts of output growth do not fully contain past information in residential investment growth and consumers' assessments of business conditions combined. We proceed by describing both the data and alternative output growth forecasts in Section 2. Section 3 presents both the methodology and forecast evaluation test results. Section 4 concludes.

2. Data and forecasts

Our study includes four sets of output growth forecasts. The first one is the Greenbook forecasts produced by the research staff at the Federal Reserve Board of Governors and is available on the Federal Reserve Bank of Philadelphia website. These forecasts in addition to the forecasts of other major macroeconomic variables are presented to the Federal Open Market Committee (FOMC) prior to each regular meeting. With the FOMC meetings occurring twice each quarter, there exist two sets of Federal Reserve forecasts. The first set is made close to the middle of the quarter and the second one is made in the last month of the quarter. In this study, we utilize the Federal Reserve forecasts made in the last month of the quarter. In addition, the Greenbook forecasts are released to the public with a five-year lag and are currently available up to the fourth quarter of 2010. In examining the one- through four-quarter-ahead forecasts, we focus on the first quarter of 1988 through the fourth quarter of 2010. As such, the sample periods for the one-, two-, three-, and four-quarter-ahead forecasts are, respectively, 1988.2–2011.1, 1988.3–2011.2, 1988.4–2011.3, and 1989.1–2011.4.

The second set of forecasts is from the Blue Chip monthly survey of private forecasters. Utilizing the individual responses, Blue Chip calculates and publishes the consensus (mean) forecasts in *Blue Chip Financial Forecasts* around the beginning of the month.[3] Given that the survey is conducted monthly, there exist three sets of forecasts for each quarter. For comparability with the Federal Reserve, we utilize the one-, two-, three-, and four-quarter-ahead Blue Chip forecasts of output growth made in the third month of the quarter.

The third set of forecasts is from a univariate AR model. In order for this benchmark to be comparable to the Federal Reserve and Blue Chip forecasts, we make use of real time data on real output available on the Federal Reserve Bank of Philadelphia website. More specifically, we utilize the data for 1967.1–1987.4 (available in the third month of 1987.4) to estimate the sample autocorrelation and partial autocorrelation function of output growth. These estimates along with the Akaike information criterion (AIC) help us select an AR(2) model with the results reported in Panel A of Table 1. As can be seen, the calculated Ljung–Box Q-statistic has a p-value well above 0.10, indicating that the residual series is white noise, and thus the model is correctly specified. We employ this model to generate the univariate AR forecasts of output growth as follows. Utilizing the 1967.1–1987.4 parameter estimates in Table 1, we generate the AR forecasts for 1988.1–1989.1. The forecast values for 1988.2, 1988.3, 1988.4, and 1989.1 correspond, respectively, to the one-, two-, three-, and four-quarter-

[2] Kydland, Rupert, and Šustek (2014) note that US investment-output dynamics cannot be generalized for other developed countries except Canada.
[3] The historical Blue Chip Financial Forecasts were purchased from Aspen Publishers, Inc.

Table 1
Univariate AR and VAR estimates: 1967.1–1987.4.

Panel A: Univariate AR estimates

$$(1-0.214B-0.158B^2)Y_t=2.788,$$
$$(1.92)\quad(1.42)\quad(3.86)$$

$R^2=0.09$, Q-statistic p-value = 0.695, inverted AR roots = 0.52; −0.30

Panel B: VAR estimates

$$Y_t=-0.099-0.170\,Y_{t-1}-0.012\,Y_{t-2}+0.084\,I_{t-1}+0.008\,I_{t-2}+0.027\,S_{t-1}-0.010\,S_{t-2}$$
$$(0.07)\quad(1.37)\quad(0.11)\quad(4.00)\quad(0.33)\quad(1.73)\quad(0.72)$$

$R^2=0.38$, Q-statistic p-value = 0.627,

$$I_t=8.204+0.480\,I_{t-1}+0.285\,I_{t-2}-1.241\,Y_{t-1}-1.380\,Y_{t-2}+0.057\,S_{t-1}-0.051\,S_{t-2}$$
$$(0.07)\quad(3.97)\quad(2.20)\quad(1.74)\quad(2.05)\quad(0.64)\quad(0.62)$$

$R^2=0.36$, Q-statistic p-value = 0.545,

$$S_t=45.96+0.640\,S_{t-1}+0.112\,S_{t-2}-1.629\,Y_{t-1}+0.260\,Y_{t-2}+0.571\,I_{t-1}+0.376\,I_{t-2}$$
$$(3.69)\quad(4.81)\quad(0.92)\quad(1.55)\quad(0.26)\quad(3.20)\quad(1.96)$$

$R^2=0.80$, Q-statistic p-value = 0.242.

Notes: Y, I, and S are, respectively, output growth, residential investment growth, and consumers' assessments of current and expected business conditions. B is the backward shift operator. The absolute t-values are in parentheses. The Ljung–Box Q-statistic test detects serial correlation up to the 12th order.

ahead Federal Reserve and Blue Chip forecasts made in the third month of 1988.1. We utilize output growth data available in the third month of 1988.1 to re-estimate the AR model for 1967.1–1988.1. The updated parameter estimates are then used to generate the AR forecasts for 1988.2–1989.2. The forecasts for 1988.3, 1988.4, 1989.1, and 1989.2 correspond, respectively, to the one-, two-, three-, and four-quarter-ahead Federal Reserve and Blue Chip forecasts made in the third month of 1988.2. This procedure is repeated until the last set of forecasts is generated for 2010.4–2011.4 using the 1967.1–2010.3 parameter estimates (note that these parameter estimates are obtained using output growth data available in the last month of 2010.3). The forecasts for 2011.1, 2011.2, 2011.3, and 2011.4 correspond, respectively, to the one-, two-, three-, and four-quarter-ahead Federal Reserve and Blue Chip forecasts made in the third month of 2010.4. As such, the sample periods for the one-, two-, three-, and four-quarter-ahead forecast are, respectively, 1988.2–2011.1, 1988.3–2011.2, 1988.4–2011.3, and 1989.1–2011.4.

The fourth set of forecasts is from the VAR model which includes three variables: output growth, growth in residential investment, and consumers' assessments of current and expected business conditions. Again, in order for these forecasts to be comparable to the Federal Reserve and Blue Chip forecasts, we utilize real time data on both real output and real residential investment available on the Federal Reserve Bank of Philadelphia website.[4] The data on consumers' assessments of business conditions are available on the Michigan Surveys of Consumers (MSC) website. This survey probes consumer sentiment on personal finances, buying and business conditions, and expectations. Utilizing a nationally-representative random sample of at least 500 US households, the survey collects individual responses to approximately 50 core questions. Here, we focus on two questions. The first one asks, "Would you say that at the present time, business conditions are better or worse than they were a year ago?" Using the individual responses, the survey calculates the index values (= better − worse + 100) for consumers' assessments of current business conditions. The second question asks, "And how about a year from now, do you expect that in the country as a whole business conditions will be better, or worse than they are at present, or just about the same?" Using the individual responses, again, the survey calculates the index values (= better − worse + 100) for consumers' assessments of expected change in business conditions. In this study, we take the sum of the two indices to obtain a measure of consumers' assessments of both current and expected business conditions. As noted above, the Federal Reserve and Blue Chip forecasts are made in the last month of the quarter. Therefore, in order for the VAR forecasts to be comparable to the Federal Reserve and Blue Chip forecasts, we utilize the MSC data for the second month of the quarter.

Utilizing the data for 1967.1–1987.4 (available in the last month of 1987.4), the multi-equation AIC criterion selects a VAR(2) model. That is, the model contains three equations (one for each variable), and each equation contains a constant term with two lags of each variable in the right-hand side. Panel B of Table 1 reports the VAR parameter estimates for 1967.1–1987.4. Utilizing these estimates, we generate the VAR forecasts for 1988.1–1989.1. The forecasts for 1988.2, 1988.3, 1988.4, 1989.1 and 1989.2 correspond, respectively, to the one-, two-, three-, and four-quarter-ahead Federal Reserve, Blue Chip, and AR forecasts made in the third month of 1988.1. We utilize the data available in the third month of 1988.1 to re-estimate the VAR model for 1967.1–1988.1. The updated parameter estimates are then used to generate the forecasts for 1988.2-1989.2. The forecasts for 1988.3, 1988.4, 1989.1, and 1989.2 correspond, respectively, to the one-, two-, three-, and four-quarter-ahead Federal Reserve, Blue Chip, and AR forecasts made in the third month of 1988.2. This procedure is repeated until the last set of VAR forecasts is generated for 2010.4–2011.4 using the 1967.1–2010.3 parameter estimates

[4] Real time data on both real GDP (GNP before 1992) and real residential investment are employed to calculate the annualized quarterly percentage rate of growth (in GDP, for example) using the following formula: $Y_t=100\times(((\mathrm{RGDP}_t\div\mathrm{RGDP}_{t-1})^4)-1)$.

Table 2
Contingency table.

	Actual change	
	Upward	Downward
Correct directional predictions	n_1: $\Delta A\,(+)$ & $\Delta P\,(+)$	n_2: $\Delta A\,(-)$ & $\Delta P\,(-)$
Incorrect directional predictions	n_3: $\Delta A\,(+)$ & $\Delta P\,(-)$	n_4: $\Delta A\,(-)$ & $\Delta P\,(+)$

Notes: ΔA $(=A_{t+f}-A_{t-1})$ is the actual change, and ΔP $(=P_{t+f}-A_{t-1})$ is the predicted change. The sign (+) represents an upward move and the sign (−) represents a downward move in output growth. The numbers of correct (incorrect) sign forecasts are denoted by n_1 and n_2 (n_3 and n_4). The sample size $n=(n_1+n_2+n_3+n_4)$.

(note that these parameter estimates are obtained using the data available in the last month of 2010.3).[5] The forecasts for 2011.1, 2011.2, 2011.3, and 2011.4 correspond, respectively, to the one-, two-, three-, and four-quarter-ahead Federal Reserve, Blue Chip, and AR forecasts made in the third month of 2010.4. Again, the sample periods for the one-, two-, three-, and four-quarter-ahead forecast are, respectively, 1988.2–2011.1, 1988.3–2011.2, 1988.4–2011.3, and 1989.1–2011.4.

3. Forecast evaluation results

Our forecast evaluation focuses on answering the following five questions:

1. Are the forecasts directionally accurate?
2. Are the forecasts free of systematic bias?
3. Are the forecasts efficient?
4. Does the asymmetric information hypothesis hold?
5. Are Federal Reserve and Blue Chip forecasts more informative than VAR forecasts?

In what follows, A_{t+f} is the actual output growth rate in quarter $t+f$, and P_{t+f} is the forecast of A_{t+f} made in the last month of quarter t (the forecast horizon $f=1, 2, 3$, and 4 quarters). With A_{t-1} defined as the actual rate most recently known at the time of the forecast, $(A_{t+f}-A_{t-1})$ is the actual change and $(P_{t+f}-A_{t-1})$ is the predicted change in output growth. We measure the actual output growth rate by the revised data available 60 days after the end of the quarter.[6]

4.1. Are the forecasts directionally accurate?

To answer, we utilize the test procedure first introduced by Merton (1981) and Henriksson and Merton (1981) and later refined by Schnader and Stekler (1990), among others. Defining the actual change by $(A_{t+f}-A_{t-1})$ and the predicted change by $(P_{t+f}-A_{t-1})$, Table 2 presents the two-by-two contingency table whose elements are the numbers of correct sign predictions (n_1 and n_2) and incorrect sign predictions (n_3 and n_4). With n defined as the sample size, $\pi_{All}=(n_1+n_2)/n$ is the overall directional accuracy rate. Table 3 reports these statistics for the Federal Reserve, Blue Chip, AR, and VAR forecasts in, respectively, rows 1–4, 5–8, 9–12, and 13–16. As can be seen for these forecasts in rows 1–16, the overall directional accuracy rates (π_{All}), ranging from 0.67 to 0.75, are quite high. In testing the null hypothesis of no (directional) association between the actual and predicted changes, we use Fisher's exact test and the chi-square tests with and without Yate's continuity correction (Sinclair, Stekler, and Kitzinger, 2010). As indicated by subscript **a**, we reject the null hypothesis for every forecast in rows 1-16 and thus conclude that they are all directionally accurate.

Schnader and Stekler (1990, 1991) note that directionally accurate predictions are of value to a user. Accordingly, we maintain that the Federal Reserve and Blue Chip forecasts in addition to the AR and VAR benchmarks are all of value to a user. Diebold (1998) points out that, in many cases, a symmetric loss structure is a good proxy for the true losses of a forecaster. In particular, a forecast generated under symmetric loss is of value to a user who assigns the same loss (cost) to both incorrect upward and downward moves. In order to see if this is the case with the forecasts under examination, we further report $\pi_{Up}=n_1/(n_1+n_3)$ which is the proportion of correctly predicted upward moves, and $\pi_{Down}=n_2/(n_2+n_4)$ which is the proportion of correctly predicted downward moves. As can be seen, for each forecast in rows 1–8, 10, 11, and 13–16 of Table 3, π_{Up} (ranging from 0.67 to 0.79) is similar to π_{Down} (ranging from 0.62 to 0.78). We use the chi-square test described in Berenson, Levine, and Rindskopf (1988) to test the null hypothesis of "no asymmetry"; that is, the proportion of incorrectly predicted upward moves $(1-\pi_{Up})$ equals the proportion of incorrectly predicted downward moves $(1-\pi_{Down})$. With the p-values (in the last column of Table 3) above 0.10, we cannot reject the null hypothesis of "no asymmetry" for the forecasts in rows 1–8, 10, 11, and 13–16. This means that the one- through four-quarter-ahead Federal Reserve, Blue

[5] It is important to note that the VAR (as well as the AR) forecasts are generated dynamically. For example, in generating the four-quarter-ahead forecasts, we make use of the one-, two-, and three-quarter-ahead forecasts. In addition, since we are recursively estimating the AR and VAR models, the parameter estimates are allowed to vary over time.

[6] Current best practice uses 90 day numbers but in the past, 60 day numbers were used and the results were comparable.

Table 3
Directional forecast accuracy test results.

Row no.	f	Correct n_1	n_2	Incorrect n_3	n_4	π_{All}	π_{Up}	π_{Down}	p-value
Federal Reserve forecasts									
1	1	31	36	15	10	0.73^a	0.67	0.78	0.241
2	2	31	31	14	16	0.67^a	0.69	0.66	0.764
3	3	34	31	13	14	0.71^a	0.72	0.69	0.716
4	4	30	32	13	17	0.67^a	0.70	0.65	0.649
Blue Chip forecasts									
5	1	32	34	14	12	0.72^a	0.70	0.74	0.643
6	2	30	32	15	15	0.67^a	0.67	0.68	0.885
7	3	34	30	13	15	0.70^a	0.72	0.67	0.554
8	4	34	34	9	15	0.74^a	0.79	0.69	0.291
Univariate AR forecasts									
9	1	36	28	10	18	0.70^a	0.78	0.61	0.070
10	2	34	29	11	18	0.68^a	0.76	0.62	0.153
11	3	36	29	11	16	0.71^a	0.77	0.64	0.201
12	4	34	30	9	19	0.70^a	0.79	0.61	0.063
VAR forecasts									
13	1	34	35	12	11	0.75^a	0.74	0.76	0.810
14	2	32	33	13	14	0.71^a	0.71	0.70	0.925
15	3	34	32	13	13	0.72^a	0.72	0.71	0.896
16	4	32	33	11	16	0.71^a	0.74	0.67	0.457

Notes: See the notes in Table 2. $\pi_{All} = (n_1 + n_2)/n$ is the overall directional accuracy rate. $\pi_{Up} = n_1/(n_1 + n_3)$ is the proportion of correctly predicted upward moves. $\pi_{Down} = n_2/(n_2 + n_4)$ is the proportion of correctly predicted downward moves. Superscript **a** indicates that the p-values of Fisher's exact test and the chi-square tests with and without Yate's continuity correction are all below 0.10, leading to the rejection of the null hypothesis of no directional association between the actual and predicted changes. p-value (in the last column) is for testing the null hypothesis that the proportion of incorrectly predicted upward moves $(1 - \pi_{Up})$ equals the proportion of incorrectly predicted downward moves $(1 - \pi_{Down})$.

Chip, and VAR forecasts in addition to the two- and three-quarter-ahead AR forecasts are of value for a user who assigns the same loss (cost) to both incorrect upward and downward moves in output growth. For the one-, and four-quarter-ahead AR forecasts, however, π_{Up} is far above π_{Down}. With the test p-values below 0.10 in the last column (rows 9 and 12) of Table 3, we reject the null hypothesis of "no asymmetry" and thus conclude that these forecasts are of value to a user who assigns high (low) cost to incorrect upward (downward) moves in output growth.

4.2. Are the forecasts free of systematic bias?

To answer, we follow Holden and Peel (1990) and estimate the test equation,

$$A_{t+f} - \mathrm{P}_{\mathrm{t+f}} = \alpha + v_{t+f} \tag{1}$$

where $(A_{t+f} - P_{t+f})$ is the forecast error. Since P_{t+f} is made in quarter t, the forecast error follows an f^{th}-order moving-average process under the null hypothesis that the population mean forecast error (ME) equals zero (i.e., $\alpha = 0$). With the forecast errors generally heteroscedastic, we utilize the Newey and West (1987) procedure to correct for both heteroscedasticity and the inherent f^{th}-order serial correlation. Column 1 of Table 4 reports the OLS estimates of Eq. (1) with the correct (Newey-West) standard errors for the Federal Reserve, Blue Chip, AR, and VAR forecasts in, respectively, rows 1–4, 5–8, 9–12, and 13–16. Column 2 further reports the mean absolute forecast errors (MAE). As can be seen, these forecasts are all free of systematic bias since we cannot reject the null hypothesis that $\alpha = 0$. Consistent with this conclusion, the size of the ME estimate (ranging from 0.023 to 0.376) is small compared to the MAE (ranging from 1.500 to 1.866) for every forecast in rows 1–16.[7]

4.3. Are the forecasts efficient?

To answer, we first ask whether the forecasts are more accurate than a naïve benchmark. Column 3 of Table 4 reports Theil's U coefficient calculated as the mean squared error (MSE) of P_{t+f} divided by the MSE of the naïve forecast (which is the actual output growth rate for quarter t-1 known at the time of the forecast). As can be seen, Theil's U coefficient estimates (ranging from 0.505 to 0.729) are all below one. We use the Diebold and Mariano (1995) test to examine the null hypothesis

[7] Unlike the ME, the MAE measures the average size of forecast errors, without considering their direction. Therefore, tendency to systematically over- or under-predict results in the size of the ME to be equal (or very close) to the MAE. However, we should note that the comparison between ME and MAE cannot detect systematic bias if the errors cancel out over the business cycle. Sinclair, Joutz, and Stekler (2010) provide a test for this particular bias that we did not apply here.

Table 4
Accuracy test results of output growth forecasts.

Row no.	f	ME (1)	MAE (2)	U (3)	DM test p-value(4)
Federal Reserve forecasts					
1	1	0.153 (0.208)	1.555	0.530	0.002
2	2	−0.076 (0.287)	1.738	0.612	0.057
3	3	−0.159 (0.335)	1.862	0.643	0.071
4	4	−0.221 (0.368)	1.866	0.623	0.016
Blue Chip forecasts					
5	1	0.176 (0.203)	1.501	0.505	0.001
6	2	−0.025 (0.272)	1.612	0.537	0.026
7	3	−0.169 (0.328)	1.732	0.619	0.082
8	4	−0.215 (0.363)	1.753	0.590	0.015
Univariate AR forecasts					
9	1	−0.292 (0.243)	1.689	0.729	0.063
10	2	−0.348 (0.307)	1.756	0.673	0.141
11	3	−0.371 (0.344)	1.782	0.665	0.158
12	4	−0.376 (0.371)	1.778	0.606	0.022
VAR forecasts					
13	1	0.017 (0.213)	1.500	0.552	0.001
14	2	−0.023 (0.268)	1.650	0.558	0.026
15	3	−0.103 (0.308)	1.679	0.574	0.052
16	4	−0.143 (0.343)	1.716	0.543	0.006

Notes: The forecast error is defined as ($A_{t+f} - P_{t+f}$), where A_{t+f} is the actual output growth in quarter $t+f$, and P_{t+f} is a general notation for the Federal Reserve, Blue Chip, AR, and VAR forecasts of A_{t+f} made in the last month of quarter t (f is the forecast horizon). ME and MAE are, respectively, the mean forecast error and the mean absolute forecast error. The correct (Newey-West) standard errors are in parentheses. Theil's U coefficient is the mean squared error (MSE) of P_{t+f} divided by the MSE of the naïve forecast. The DM (Diebold-Mariano) test p-value below 0.10 indicates that the MSE of P_{t+f} is significantly below the MSE of the naïve forecast.

of equal forecast accuracy. As reported in column 4, the Diebold-Mariano (DM) p-values for the Federal Reserve and Blue Chip forecasts in rows 1–8 are all below 0.10, indicating that these forecasts produce significantly lower MSEs than the naïve forecast. The same is true for the one- and four-quarter-ahead AR forecasts in rows 9 and 12 and for the VAR forecasts in rows 13–16. With the DM p-values above 0.10 in rows 10 and 11, the two- and three-quarter-ahead AR forecasts do not produce *significantly* lower MSEs than the naïve benchmark.

Given such results, we now turn to see whether the forecasts are efficient. A forecast is said to be efficient if it contains past information in the target variable. By default, the AR forecast is efficient since the univariate AR model efficiently employs past information in output growth. With this in mind, we then examine the efficiency of the Federal Reserve, Blue Chip, and VAR forecasts by estimating,

$$(A_{t+f} - A_{t-1}) = \gamma_0 + \gamma_1(PA_{t+f} - A_{t-1}) + \gamma_2(P_{t+f} - A_{t-1}) + v_{t+f} \tag{2}$$

where PA_{t+f} is the AR forecast, and P_{t+f} is a general notation denoting the Federal Reserve, Blue Chip, and VAR forecasts. Following Fair and Shiller (1990), we note that PA_{t+f} and P_{t+f} contain similar information when the estimates of γ_1 and γ_2 are both insignificant. PA_{t+f} lacks the useful information contained in P_{t+f} when the estimate of γ_1 is negative or insignificant and the estimate of γ_2 is positive and significant; the converse is also true. PA_{t+f} and P_{t+f} contain distinct predictive information when the estimates of γ_1 and γ_2 are both positive and significant. In this case, combining the two forecasts yields a composite forecast that should be more informative than the individual forecasts (Granger and Ramanathan, 1984).

Table 5 reports the OLS estimates of Eq. (2) along with the correct standard errors for the Federal Reserve forecasts in rows 1–4, for the Blue Chip forecasts in rows 5–8, and for the VAR forecasts in rows 9–12. As can be seen in rows 1–3 and 5–7, the estimates of γ_1 are insignificant and the estimates of γ_2 are both positive and significant, indicating that the one- through three-quarter-ahead Federal Reserve and Blue Chip forecasts contain useful predictive information above and beyond that contained in the AR forecasts. The estimates of $\gamma_1$and γ_2 in row 4 are both insignificant, indicating that the four-quarter-ahead Federal Reserve and AR forecasts contain similar information. The same is true for the four-quarter-ahead Blue Chip and AR forecasts in row 8. Put together, such results indicate that the Federal Reserve and Blue Chip forecasts of output growth are all efficient. The results in rows 9–12 further indicate that the VAR forecasts are also efficient since the estimates of γ_1 are negative but the estimates of γ_2 are both positive and significant. More specifically, such findings indicate that the VAR forecasts embody useful predictive information above and beyond that contained in the AR forecasts which, in turn, allow us to conclude that past information on growth in residential investment and consumers' assessments of business conditions is useful in predicting output growth.

Table 5
Efficiency test results.

Row no.		EQ2: $(A_{t+f} - A_{t-1}) = \gamma_0 + \gamma_1 (PA_{t+f} - A_{t-1}) + \gamma_2 (P_{t+f} - A_{t-1}) + v_{t+f}$			
	f	γ_0	γ_1	γ_2	R^2
Federal Reserve forecasts					
1	1	0.080 (0.196)	0.157 (0.206)	0.810^b(0.171)	0.48
2	2	−0.156 (0.247)	0.315 (0.323)	0.637^b(0.261)	0.41
3	3	−0.216 (0.291)	0.361 (0.486)	0.536^b(0.321)	0.39
4	4	−0.312 (0.339)	0.622 (0.395)	0.356 (0.356)	0.42
Blue Chip forecasts					
5	1	0.379 (0.209)	−0.441 (0.329)	1.418^b(0.300)	0.52
6	2	0.156 (0.238)	−0.549 (0.500)	1.495^b(0.446)	0.48
7	3	−0.125 (0.265)	−0.090 (0.721)	0.958^b(0.571)	0.39
8	4	−0.250 (0.314)	0.293 (0.682)	0.659 (0.646)	0.42
VAR forecasts					
9	1	0.115 (0.189)	−0.313 (0.362)	1.279^b (0.314)	0.46
10	2	0.086 (0.259)	−0.325 (0.573)	1.273^b (0.512)	0.45
11	3	0.202 (0.290)	−1.090 (0.838)	1.989^b (0.758)	0.46
12	4	0.290 (0.324)	−1.810 (1.203)	2.745^b (1.160)	0.49

Notes: PA_{t+f} is the univariate AR forecast, and P_{t+f} is a general notation for the Federal Reserve, Blue Chip, and VAR forecasts. The correct (Newey-West) standard errors are in parentheses. Superscript **b** indicates significance at the 10% or lower level.

Table 6
Encompassing test results.

Row no.		EQ3: $(A_{t+f} - A_{t-1}) = \delta_0 + \delta_1 (PF_{t+f} - A_{t-1}) + \delta_2 (PB_{t+f} - A_{t-1}) + v_{t+f}$			
	f	δ_0	δ_1	δ_2	R^2
Federal Reserve vs. Blue Chip forecasts					
1	1	0.174 (0.202)	0.392^b(0.220)	0.668^b(0.248)	0.52
2	2	−0.027 (0.278)	0.040 (0.245)	0.967^b(0.287)	0.46
3	3	−0.142 (0.326)	0.293 (0.426)	0.586 (0.510)	0.40
4	4	−0.198 (0.374)	0.130 (0.439)	0.797 (0.461)	0.42
		EQ4: $(A_{t+f} - A_{t-1}) = \eta_0 + \eta_1 (PF_{t+f} - A_{t-1}) + \eta_2 (PR_{t+f} - A_{t-1}) + v_{t+f}$			
	f	η_0	η_1	η_2	R^2
Federal Reserve vs. VAR forecasts					
5	1	0.090 (0.195)	0.568^b (0.168)	0.557^b(0.179)	0.55
6	2	−0.046 (0.267)	0.381^b (0.175)	0.661^b(0.174)	0.47
7	3	−0.111 (0.317)	0.230 (0.232)	0.734^b(0.302)	0.44
8	4	−0.149 (0.361)	0.052 (0.289)	0.959^b(0.303)	0.46
		EQ5: $(A_{t+f} - A_{t-1}) = \omega_0 + \omega_1 (PB_{t+f} - A_{t-1}) + \omega_2 (PR_{t+f} - A_{t-1}) + v_{t+f}$			
	f	ω_0	ω_1	ω_2	R^2
Blue Chip vs. VAR forecasts					
9	1	0.131 (0.196)	0.751^b(0.248)	0.391^b(0.237)	0.52
10	2	−0.027 (0.267)	0.641^b (0.286)	0.405^b(0.254)	0.48
11	3	−0.100 (0.312)	0.043 (0.375)	0.913^b(0.400)	0.43
12	4	−0.121 (0.348)	−0.334 (0.589)	1.348^b(0.604)	0.46

Notes: PF_{t+f}, PB_{t+f}, and PR_{t+f} denote, respectively, the Federal Reserve, Blue Chip, and VAR forecasts. The correct (Newey-West) standard errors are in parentheses. Superscript **b** indicates significance at the 10% or lower level.

4.4. Does the asymmetric information hypothesis hold?

In examining this question, we follow Fair and Shiller (1990) and estimate,

$$(A_{t+f} - A_{t-1}) = \delta_0 + \delta_1(PF_{t+f} - A_{t-1}) + \delta_2(PB_{t+f} - A_{t-1}) + v_{t+f} \tag{3}$$

where PF_{t+f} is the Federal Reserve forecast, and PB_{t+f} is the Blue Chip forecast. Table 6 reports the OLS estimates of Eq. (3) along with the correct standard errors in rows 1–4. As can be seen, the estimates of δ_1 and δ_2 in row 1 are both positive and significant, indicating that the one-quarter-ahead Federal Reserve and Blue Chip forecasts contain distinct predictive information. In this case, combining the two forecasts yields a composite forecast that should be more informative than the individual forecasts. With the estimate of δ_1 insignificant and the estimate of δ_2 positive and significant in row 2, the two-quarter-ahead Blue Chip forecast contains useful predictive information above and beyond that contained in the Federal Reserve forecasts. The results in rows 3 and 4 further indicate that the three- and four-quarter-ahead Federal Reserve and

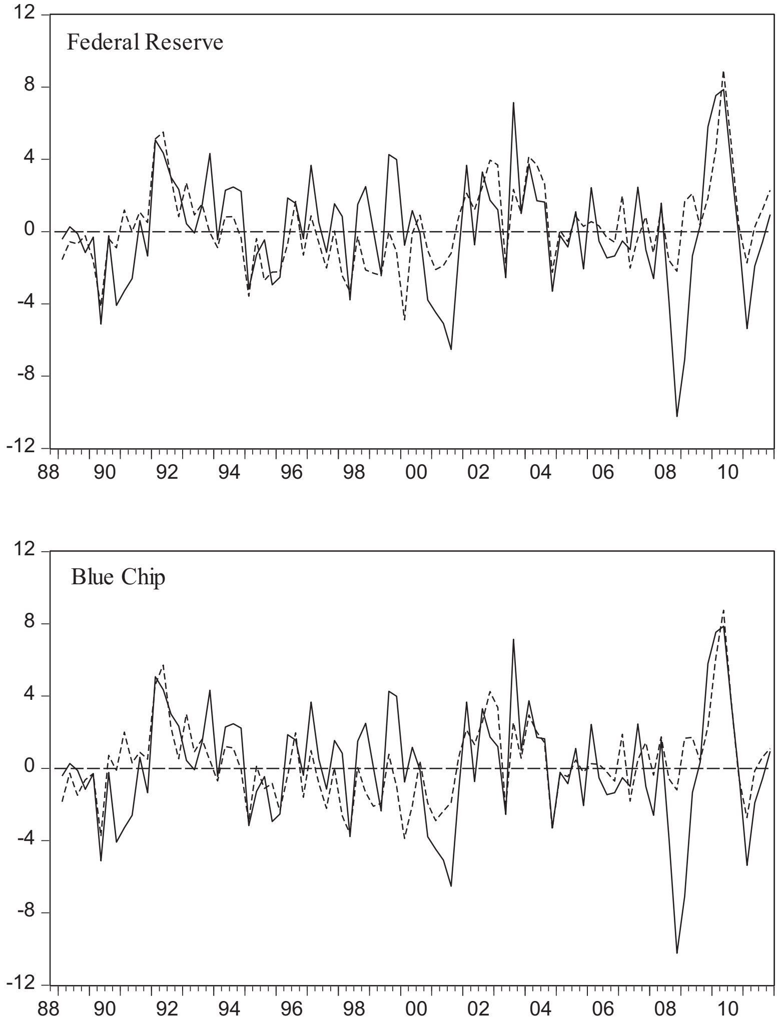

Fig. 1. Actual change (solid line) vs. four-quarter-ahead predicted change (dotted line) in output growth: 1989.1–2011.4.

Blue Chip forecasts contain similar information since the estimates of δ_1 and δ_2 are both insignificant. As depicted in Fig. 1, the four-quarter-ahead Federal Reserve and Blue Chip predicted change series follows the actual change series with a very similar pattern. Put together, our results fail to support the asymmetric information hypothesis that the Federal Reserve has useful information about the state of the economy that is not known by the private sector.

4.5. *Are Federal Reserve and Blue Chip forecasts more informative than VAR forecasts?*

In answering, again, we follow Fair and Shiller (1990) and estimate the test equations,

$$(A_{t+f} - A_{t-1}) = \eta_0 + \eta_1(PF_{t+f} - A_{t-1}) + \eta_2(PR_{t+f} - A_{t-1}) + \nu_{t+f} \quad (4)$$

$$(A_{t+f} - A_{t-1}) = \omega_0 + \omega_1(PB_{t+f} - A_{t-1}) + \omega_2(PR_{t+f} - A_{t-1}) + \nu_{t+f} \quad (5)$$

where PR_{t+f} is the VAR forecast. Table 6 reports the OLS estimates of Eq. (4) along with the correct standard errors in rows 5–8. As can be seen in rows 5 and 6, the estimates of η_1 and η_2 are both positive and significant, indicating that the one- and two-quarter-ahead Federal Reserve and VAR forecasts contain distinct predictive information. In this case, combining the two forecasts yields a composite forecast that should be more informative than the individual forecasts. With the estimates of η_1 insignificant and the estimates of η_2 positive and significant in rows 7 and 8, we conclude that the three- and four-quarter-ahead Federal Reserve forecasts lack the predictive information in the VAR forecasts.

Table 6 further reports the OLS estimates of Eq. (5) along with the correct standard errors in rows 9–12. As can be seen in rows 9 and 10, the estimates of ω_1 and ω_2 are both positive and significant, indicating that the one- and two-quarter-ahead Blue Chip and VAR forecasts contain distinct predictive information. Again, in this case, combining the two forecasts yields a composite forecast that should be more informative than the individual forecasts. With the estimates of ω_1 insignificant and the estimates of ω_2 positive and significant in rows 11 and 12, we conclude that the three- and four-quarter-ahead Blue Chip forecasts lack predictive information in the VAR forecasts.

5. Conclusions

A good deal of literature concerns modeling and forecasting future growth in real output (Chauvet and Potter, 2013). In particular, decisions are guided by forecasts, and good forecasts are important in helping policymakers formulate successful economic policy. In this study, we compare the predictive content of both the Federal Reserve and Blue Chip forecasts of output growth using comparable AR and VAR forecasts. The univariate AR forecast contains past information in output growth, and the VAR forecast contains past information in output growth, growth in residential investment, and consumers' assessments of business conditions known at the time of the forecast. We find that the VAR forecasts are more informative than the AR forecasts, indicating that growth in residential investment and consumers' assessments of business conditions together are useful in predicting output growth.

Some advocate that the Fed should take preemptive measures to prevent major economic and financial crises. However, effective preemptive measures require timely forecasts of major economic indicators including output growth. Our findings that the Federal Reserve (including private) forecasts of output growth generally lack past information in residential investment growth and consumers' assessments of business conditions is not encouraging. Given the strong link between housing market activity and business cycles (Leamer, 2007), our study suggests that Federal Reserve and private forecasters need to be more mindful of housing market activities. A similar suggestion follows with regard to consumer survey data due to their significant predictive power for output growth.

Further, our test results do not support the asymmetric information hypothesis that the Federal Reserve has useful information about the state of the economy that is not known by the private sector. In particular, the one-quarter-ahead Federal Reserve and Blue Chip forecasts contain distinct predictive information, the two-quarter-ahead Blue Chip is more informative than the Federal Reserve forecast, and the three- and four-quarter-ahead Federal Reserve and Blue Chip forecasts contain similar information. Failure to support the asymmetric information hypothesis may not be of concern, since the FOMC members who vote on policy are mindful of the private forecasts (Baghestani 2008). As such, we re-emphasize our findings that the VAR forecasts are generally more informative than the Federal Reserve and private forecasts and the suggestion that, in forecasting output growth, the Fed should make greater use of the information in both housing market indicators and consumers' assessments of current and future business conditions.

Acknowledgements

The authors gratefully acknowledge helpful comments and suggestions of Editor Elyas Elyasiani and two anonymous referees.

References

Baghestani, H. (2008). Federal Reserve vs. private information: Who is the best unemployment rate predictor? *Journal of Policy Modeling*, *29*, 101–110.
Baghestani, H. (2011). Federal Reserve and private forecasts of growth in investment. *Journal of Economics and Business*, *63*, 290–305.
Batchelor, R., & Dua, P. (1998). Improving macro-economic forecasts: The role of consumer confidence. *International Journal of Forecasting*, *14*, 71–81.
Berenson, M. L., Levine, D. M., & Rindskopf, D. (1988). *Applied Statistics: A First Course*. New York: Prentice Hall.
Bram, J., & Ludvigson, S. (1998). Does consumer confidence forecast household expenditure? A sentiment index horse race. *Economic Policy Review*, *4*, 59–78.
Carroll, C. D., Fuhrer, J. C., & Wilcox, D. W. (1994). Does consumer sentiment forecast household spending? If so, why? *American Economic Review*, *84*, 1397–1408.
Chauvet, M., & Potter, S. (2013). Forecasting output. In G. Elliott, & A. Timmermann (Eds.), *Handbook of Economic Forecasting* (Vol. 2). Elsevier.
Christiansen, C., Eriksen, J. N., & Møller, S. V. (2014). Forecasting US recessions: The role of sentiment. *Journal of Banking & Finance*, *49*, 459–468.
Coulson, N. E., & Kim, M. S. (2000). Residential investment, non-residential investment and GDP. *Real Estate Economics*, *28*, 233–247.
Croushore, D. (2005). Do consumer confidence indexes help forecast consumer spending in real time? *North American Journal of Economics and Finance*, *16*, 435–450.
Dees, S., & Brinca, P. S. (2013). Consumer confidence as a predictor of consumption spending: Evidence for the United States and the Euro area. *International Economics*, *134*, 1–14.
Diebold, F. X. (1998). *Elements of Forecasting*. Cincinnati, Ohio: South-Western College Publishing.
Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*, 253–263.
Fair, R. C., & Shiller, R. J. (1990). Comparing information in forecasts from econometric models. *American Economic Review*, *80*, 375–389.
Granger, C. W., & Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, *3*, 197–204.
Garner, C. A. (1991). Forecasting consumer spending: Should economists pay attention to consumer confidence surveys? *Economic Review*, *76*, 57–71. Federal Reserve Bank of Kansas City.
Gavin, W. T., & Mandal, R. J. (2001). Forecasting inflation and growth: Do private forecasts match those of policymakers? *Business Economics*, *36*, 13–20.
Green, R. (1997). Follow the leader: How changes in residential and non-residential investment predict changes in GDP. *Real Estate Economics*, *25*, 253–270.
Henriksson, R. D., & Merton, R. C. (1981). On market timing and investment performance. II. Statistical procedures for evaluation forecasting skills. *Journal of Business*, *54*, 513–533.
Holden, K., & Peel, D. A. (1990). On testing for unbiasedness and efficiency of forecasts. *The Manchester School*, *58*, 120–127.
Kydland, F.E., P. Rupert, R. Šustek (2014). Housing dynamics over business cycles: An international perspective. Center for Macroeconomics Paper No. CFM-DP 2014-23.
Leamer, E. E. (2007). Housing is the business cycle. *Proceedings, Reserve Bank of Kansas City*, 149–233.

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.

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*, 703–708.

Österholma, P. (2014). Survey data and short-term forecasts of Swedish GDP growth. *Applied Economics Letters*, *21*, 135–139.

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

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

Schnader, M. H., & Stekler, H. O. (1991). Evaluating predictions of change: An application to inflation forecasts. *Applied Financial Economics*, *1*, 135–137.

Sims, C. (2002). The role of models and probabilities in the monetary policy process. *Brookings Papers on Economic Activity*, 1–62.

Sinclair, T. M., Stekler, H. O., & Kitzinger, L. (2010). Directional forecasts of GDP and inflation: A joint evaluation with an application to Federal Reserve predictions. *Applied Economics*, *42*, 2289–2297.

Sinclair, T. M., Joutz, F., & Stekler, H. O. (2010). Can the Fed predict the state of the economy? *Economics Letters*, *108*, 28–32.