International Journal of Forecasting 32 (2016) 313–323

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Revisiting the relative forecast performances of Fed staff and private forecasters: A dynamic approach

Makram El-Shagi [a,b], Sebastian Giesen [c,*], Alexander Jung [d]

[a] School of Economics, Henan University, PR China
[b] Halle Institute for Economic Research, Germany
[c] Deutsche Bundesbank, Germany
[d] European Central Bank, Germany

ARTICLE INFO

Keywords:
Forecast performance
Forecast stability
Greenbook forecasts
Survey of Professional Forecasters

ABSTRACT

This paper aims to extend the findings of Romer and Romer (2000) to a setup where the time variation of (relative) forecast performances is addressed in much greater detail. We show that the relative forecast performances of Fed staff and private forecasters are not stable in the presence of large macroeconomic shocks such as the Great Moderation and the oil price shocks of the 1970s. Furthermore, we show that the predictive ability of the staff outperforms that of private forecasters in the presence of specific factors, such as an increased uncertainty in the economy and the staff's better knowledge of the Fed's future interest rate.

1. Introduction

Today, most central bank watchers invest vast resources in the production of good forecasts of inflation and output. Why is this? It is because it helps them to assess the monetary policy stance better in real time, and to form expectations about the likely future interest rate path. Can private forecasters learn something from central bank forecasts of these key macroeconomic variables? Romer and Romer (2000) demonstrated that the Federal Reserve's (Fed) Greenbook forecasts outperform private forecasts of output and inflation in the United States, but have relationships changed since then, or has the US economy become more predictable, as was suggested by Tulip (2009)? Moreover, the presence of information cascades (see Bikhchandani, Hirshleifer, & Welch, 2008) appears to have contributed to a better sharing of information and a reduction in the dispersion of private forecasts.

For at least three reasons, it is questionable whether the superiority of staff forecasts still holds today. First, the Fed and other main central banks have achieved a high level of transparency, thereby reducing their relative information advantage, with the aim of enhancing the effectiveness of monetary policy (see Woodford, 2005). In this context, only the Fed's Summary of Economic Projections is published in real time, while its Greenbook forecasts are released with a lag of five years. Second, given several large macroeconomic shocks (the Great Moderation, oil price shocks, and financial crises) which contributed to changes in the volatility patterns of macroeconomic time series, it is conceivable that the relative forecast performances of staff and private forecasters may have changed. However, because both groups of forecasters have been subject to profound uncertainty in the presence of these shocks, it could be that their performances are different. Third, the superiority of Greenbook forecasts is at odds with the findings of another paper by Romer and Romer (2008), which suggests that

* Corresponding author.
E-mail addresses: Makram.ElShagi@gmail.com (M. El-Shagi), Sebastian.Giesen@bundesbank.de (S. Giesen), Alexander.Jung@ecb.europa.eu (A. Jung).

http://dx.doi.org/10.1016/j.ijforecast.2015.05.006
0169-2070/

the forecasts made by the Federal Reserve Open Market Committee (FOMC) are less informative than the Greenbook forecasts. These forecasts by FOMC policy-makers are informed by the internal Greenbook forecasts, and should not differ fundamentally from the latter. Ellison and Sargent (2012) have challenged this view in their defence of the FOMC policy-makers' forecasts.

Another strand of the literature has argued that monetary policy should be rule-based, in order to deal with economic uncertainty. This could be viewed as suggesting that forecasts are less relevant for policy-makers. Does the Greenbook's forecast accuracy matter for US monetary policy? The answer depends strongly on the persistence of the macroeconomic processes. If high levels of inflation or output today signal high levels of those variables in the future, for example, a distinction between the current and forecast values may not matter much (see Bernanke, 2010). The distinction between current and forecast levels of inflation has become increasingly important, as has been shown by several oil price shocks, which have led to increases in the overall inflation. Moreover, as the FOMC transcripts make clear, Fed policymakers use forward-looking Taylor rules as a benchmark for their internal discussions. In fact, the FOMC uses a suite of these rules to check for robustness. In order to be a meaningful benchmark for policy, the Taylor rule needs to have good forecasts as inputs.

The aim of this paper is to extend the findings of Romer and Romer to a setup in which the time variation of (relative) forecast performances is addressed in much greater detail. To address this issue, the paper uses the latest available econometric techniques, such as the fluctuation test of Giacomini and Rossi (2010) and the test of conditional predictive ability by Giacomini and White (2006), as well as rolling window Mincer–Zarnowitz regressions and a battery of fixed event Nordhaus (1987) tests. Using an extended sample (1968–2006), we confirm the results of Romer and Romer (2000) that the Greenbook forecasts for inflation and output are more accurate than those from private forecasters. However, we find that this superiority is driven mostly by central bank staff performing much better in a few periods, which usually coincide with times of great macroeconomic distress. In particular, our econometric tests show that relatively better forecast performances by staff are observed when there is an increased uncertainty. The staff's greater knowledge about the Fed's future interest rate path also plays an important role in this respect.

The paper is organized as follows. Section 2 briefly reviews the literature. Section 3 explains the data used in this study and revisits the traditional battery of full sample forecast performance tests for the full extended sample. Section 4 analyses the dynamics of forecast performances, testing the changes in forecast rationality and relative forecast performances over time, and – most importantly – identifying the driving factors that underlie the fluctuations in relative forecast performances. Section 5 concludes.

2. A brief review of the literature

Greenbook forecasts are thought to provide the FOMC with an information advantage relative to private forecasters because (see Romer & Romer, 2000), first, Greenbook forecasts are more accurate than private forecasts, that is, they have lower root mean square errors (RMSE); and, second, private sector forecasts have little or no additional explanatory power for inflation, relative to the Fed's Greenbook forecast. The relative performances of private sector and staff forecasts have been the subject of a series of empirical studies. Due to data availability limitations, most of the studies examining this issue have considered the US. For various different samples, ranging from the late 1960s to the mid-1990s, several studies investigating the forecast accuracy have supported the finding on the Fed's information advantage (see D'Agostino & Whelan, 2008; Gavin & Mandal, 2003; Peek, Rosengren, & Tootell, 2003; Reifschneider & Tulip, 2007; Sims, 2002). However, a more recent study by Gamber and Smith (2009) found that the gap between the Greenbook forecasts and private forecasts has narrowed considerably since the mid-1980s, especially since 1994.

Why might the gap between staff and private forecasts have narrowed? One of the main reasons is the high level of transparency that has been achieved by main central banks over the last decade. They have deliberately reduced their relative information advantage in order to enhance the effectiveness of monetary policy (see Woodford, 2005). Swanson (2006) suggests that increases in Fed transparency since the late 1980s have been instrumental in enabling both US financial markets and the private sector to forecast the federal funds rate at horizons of several months. Several authors (see Blattner, Catenaro, Ehrmann, Strauch, & Turunen, 2008; Brand, Buncic, & Turunen, 2010; Lange, Sack, & Whitesell, 2003) find an increased predictability of the FOMC decisions due to the improved transparency. In regard to inflation forecasts by central bank staff (and possibly also output forecasts, if they are made consistent with the inflation forecast), a further argument is that the price stability goal has become more important over time. Thus, central banks could have an incentive to align their forecasts with their numerical inflation target at the policy horizon. Such a behaviour could result in a deterioration of the forecast accuracy, both in absolute terms and relative to other forecasters (see e.g. Jung, 2013).

The Fed Greenbook staff forecasts for each FOMC meeting are published only with a lag of about five years, whereas, in principle, the FOMC forecasts made by policymakers are available to market observers in real-time. Since the staff forecasts are not available to the public when they are assessing the Fed's monetary policy stance and the economic outlook, an interesting question that has been examined in other papers is whether the Greenbook forecasts are superior to the FOMC forecasts. Romer and Romer (2008) find that the FOMC forecasts do not provide useful information relative to the Greenbook forecast, even though the FOMC members know the staff forecast when making their individual forecasts. The accuracy of the FOMC forecasts could be influenced negatively by specific factors. Strategic motives by individual members and a non-harmonised interest rate assumption (see McCracken, 2010; Tillmann, 2011) are examples of such factors. However, the study by Ellison and Sargent (2012) suggests that

the idea of the inferiority of FOMC forecasts is at odds with evidence showing that the differences between the FOMC and Greenbook forecasts are very small. Recent work by Sheng (2015) provides evidence that FOMC members' forecasts are superior to those of private forecasters. In addition, an empirical analysis by Nunes (2013) suggests that FOMC forecasts are a good proxy for Greenbook forecasts in real time, but that they are different. This observation can be explained by the fact that FOMC members and Fed staff may use different models, judgement and objectives (see Bhattacharjee & Gelain, 2011).

3. Data and full sample evidence

3.1. Data

This paper uses quarterly forecasts for inflation (GDP deflator) and (real) output from the Greenbook forecasts and the Survey of Professional Forecasters (SPF) for the US. It includes Greenbook forecasts and outcomes for the sample 1968Q4 to 2006Q4 from the real-time database of the Federal Reserve Bank of Philadelphia.[1] For a more detailed analysis of GDP and inflation forecasts, the database also includes Greenbook forecasts and outcomes of the GDP components (i.e. real consumption, real fixed business investment, real residential investment, real federal government consumption, and real local and state government consumption), as well as those for nominal GDP and CPI inflation. The Greenbook projections are prepared independently by the research staff at the Board of Governors for each FOMC meeting, without interference from the Board. Greenbook forecasts are generally available for five or six quarters into the future, though the horizon of the forecasts varies over time and with the date of the FOMC meeting. They generally report forecasts in terms of real GDP growth, but the data reported before 1992 are for real GNP. Likewise, for inflation, longer series are typically reported for the GDP (GNP) deflator. Other measures for inflation are available at somewhat shorter horizons, namely the CPI since 1979 and the PCE since 2000. These measures have been used more prominently in policy debates.

The SPF is the oldest quarterly survey of macroeconomic forecasts in the US. It was initially conducted by the American Statistical Association and the National Bureau of Economic Research. The Bank of Philadelphia then took over the survey in 1990, and transformed it into a real-time survey as of the third quarter of 1990. Forecasts for core CPI inflation, PCE inflation, and core PCE inflation were only added as late as the first quarter of 2007. We therefore use the GDP (GNP) deflator (and also report tests for CPI inflation where meaningful). For real GDP (GNP), the observations are fully comparable with the Greenbook forecasts. In this context, the present analysis uses the mean survey responses, which are based on the implied forecasts for each panellist (for details, see the website of the Federal Reserve of Philadelphia). These forecasts are available for the long time series, and their calculation takes into account the fact that the sample composition has changed and that there have been no forecasts reported by individual forecasters for some periods. Using the mean also takes into consideration possible pooling advantages among private forecasters, since the mean forecasts give little weight to extreme values.

Staff forecasters have a slight timing advantage over the SPF, and private forecasts have no knowledge of the Greenbook forecasts until five years later. We use the Philadelphia Fed's Greenbook dataset, which matches the timing of the SPF forecasts. The Greenbook forecasts are released to the FOMC members prior to each meeting (the time of month when the forecast is made also varies, because the date of the FOMC meeting varies). For the purposes of this study, we use the observations that became available in March, June, September and December. For the SPF, the timing is such that new observations are released to the public by the middle or end of February, May, August and November.[2]

3.2. Full sample tests of unbiasedness and efficiency

We employ two full sample tests for the unbiasedness of the forecasts. First, we employ a standard Mincer–Zarnowitz regression (which is applied frequently, see e.g. Romer & Romer, 2000; Rossi & Sekhposyan, in press). By regressing the outcome at time $t + h$, $A_{h,t}$, on the h-period-ahead forecast made at time t, $F_{h,t}$, using the test equation

$$A_{h,t} = \alpha + \beta \hat{F}_{h,t} + \varepsilon_{h,t}, \tag{1}$$

it is possible to assess unbiasedness by jointly testing the null hypothesis that the constant term α equals zero (i.e. that there is no level shift) and the slope parameter β equals 1 (i.e. that a change in the forecast does reflect a change in the outcome of the same size on average).

Second, we use the test proposed by Holden and Peel (1990):

$$A_{h,t} - \widehat{F}_{h,t} = \alpha + \eta_{h,t}, \tag{2}$$

which tests the null that the constant in the equation equals zero, where η is a MA($h - 1$) term to account for the autocorrelation of forecast errors that is induced by overlapping forecast horizons.[3] While Holden and Peel (1990) have shown that the traditional Mincer–Zarnowitz regression generally overrejects, the MA term causes a small sample bias, thus distorting the results for samples of our size. In our sample, neither test rejects unbiasedness for the inflation or GDP growth forecasts of either staff or private forecasters; see Table 1.

[1] All of the results reported in this paper are based on real-time data, i.e., first releases. For the US, GDP data are subject to deep revisions, and this could have a bearing on the results. In fact, when using final vintages, the results for GDP are not fully robust, whereas the results for inflation, for which revisions have been less substantial, are robust.

[2] For details, see the Federal Reserve of Philadelphia's documentation: http://www.phil.frb.org/research-and-data/real-time-center/survey-of-professional-forecasters/spf-documentation.pdf.

[3] We account for this problem with the Mincer–Zarnowitz regression using HAC standard errors.

Table 1
Tests for forecast rationality.

Horizon	α		β		R^2		Wald test		HP test		KR test		BP test	
	GB	SPF	GB	SPF	GB	SPF	GB	GB	SPF	SPF	GB	SPF	GB	SPF
Inflation (GDP deflator)														
Nowcast	0.07 (0.22)	−0.30 (0.21)	0.96 (0.06)	1.05 (0.06)	0.67	0.82	0.85	0.30	0.17	0.51	0.51	0.59	80Q4	81Q1
1 quarter ahead	0.06 (0.29)	−0.21 (0.31)	1.00 (0.08)	1.05 (0.09)	0.75	0.68	0.90	0.78	0.65	0.91	0	0	80Q4	80Q4
2 quarters ahead	0.08 (0.30)	−0.20 (0.36)	1.01 (0.09)	1.04 (0.10)	0.68	0.59	0.80	0.86	0.56	0.91	0	0	80Q3	80Q4
3 quarters ahead	0.04 (0.33)	−0.16 (0.47)	1.01 (0.10)	1.03 (0.13)	0.65	0.51	0.92	0.94	0.74	0.87	0	0	80Q2	81Q2
4 quarters ahead	0.01 (0.38)	−0.02 (0.55)	1.01 (0.12)	0.99 (0.14)	0.61	0.43	0.99	0.98	0.86	0.89	0	0	80Q1	81Q1
Real output (GDP)														
Nowcast	0.24 (0.20)	−0.11 (0.23)	0.96 (0.05)	1.12 (0.07)	0.64	0.64	0.49	0.10	0.39	0.22	0.78	0.78	79Q3	79Q3
1 quarter ahead	0.30 (0.34)	−0.21 (0.39)	0.83 (0.10)	1.03 (0.12)	0.33	0.33	0.16	0.83	0.53	0.65	0.19	0.14	79Q2	79Q2
2 quarters ahead	0.10 (0.44)	−0.32 (0.39)	0.85 (0.13)	1.00 (0.17)	0.19	0.19	0.20	0.37	0.32	0.31	0	0	80Q1	79Q1
3 quarters ahead	0.92 (0.54)	0.33 (0.81)	0.56 (0.16)	0.72 (0.22)	0.06	0.06	0.01	0.07	0.23	0.15	0.71	0.22	79Q4	79Q4
4 quarters ahead	0.83 (0.61)	0.05 (0.87)	0.63 (0.19)	0.79 (0.26)	0.06	0.06	0.09	0.04	0.38	0.14	0.56	0.87	92Q4	82Q1

Notes: Sample: 1968Q4 to 2006Q4 (about 150 observations were included). GB: Greenbook forecasts. SPF: Survey of Professional Forecasters. HAC standard errors are in brackets. HP test: test by Holden and Peel (1990); KR test: test by Keane and Runkle (1990); BP test: Andrews et al. (1996) breakpoint test.

In addition, we also test for (weak) efficiency by estimating an augmented form of Eq. (1), as proposed by Keane and Runkle (1990):

$$A_{h,t} = \alpha + \beta \hat{F}_{h,t} + \gamma (A_{h,t-1} - \hat{F}_{h,t-1}) + \nu_{h,t}, \tag{3}$$

where v follows a moving average (MA) process of appropriate order. The null hypothesis of weak efficiency is H_0: $\gamma = 0$, i.e. forecast errors have no predictive power for the dependent variable.

For output, the test results support weak efficiency for both the Fed staff and SPF forecasts. For inflation, though, the tests reject efficiency for both types of forecasters at all horizons, except for the nowcast. Uncertainty concerning the timing of future shifts in inflation is one factor that may be responsible for autocorrelated forecast errors. Efficiency is not rejected for nowcasts, since such uncertainty mostly affects forecasts that are based on structural (or structurally inspired) models. At the same time, expectations of an upcoming shift in inflation may turn out to be well founded, but with a different timing. Hence, the detected "inefficiency" of inflation forecasts at longer horizons should not be interpreted as an outright rejection of their rationality.

3.3. Encompassing tests: testing for additional information

It is widely believed that central banks have access to superior information when assessing the economic situation and the future course of monetary policy. If this is the case, could private forecasters improve their forecasts by learning from the central bank staff forecast? To test whether staff forecasts encompass private forecasts or vice versa, or whether both contain additional information, we regress the outcome on both forecasts using the equation:

$$A_{h,t} = \delta + \gamma^P \hat{F}^P_{h,t} + \gamma^S \hat{F}^S_{h,t} + \nu_{h,t}, \tag{4}$$

where $\hat{F}_{h,t}$ is the corresponding h-step-ahead forecast from the central bank staff (superscript S) or the private forecaster (superscript P). A significantly positive value of γ indicates that the respective forecast provides information that is not contained in the other forecast.

When assessing the possible impact of the timing on the relative forecast performances, we repeat the tests from Eq. (4) with an important modification. We put central bank staff at a timing disadvantage of one quarter, i.e., we use their forecasts from the previous quarter, and check whether these forecasts are still informative for private forecasters. We estimate:

$$A_{h,t} = \delta + \gamma^P \hat{F}^P_{h,t} + \gamma^S \hat{F}^S_{h+1,t-1} + \nu_{h,t}. \tag{5}$$

Tables 2 and 3 report estimation results for Eqs. (4) and (5). We confirm the findings of Romer and Romer (2000) on the potential usefulness of Greenbook forecasts for the private sector. Our results indicate that for an extended

Table 2
Encompassing test.

Horizon	δ	γ^P	γ^S	Adj. R^2
Inflation (GDP deflator)				
Nowcast	−0.11	0.32***	0.69***	0.85
	(0.19)	(0.11)	(0.13)	
1 quarter ahead	0.11	−0.09	1.08***	0.75
	(0.33)	(0.17)	(0.16)	
2 quarters ahead	0.26	−0.26	1.23***	0.68
	(0.36)	(0.35)	(0.36)	
3 quarters ahead	0.21	−0.21	1.19**	0.65
	(0.47)	(0.35)	(0.32)	
4 quarters ahead	0.58	−0.77*	1.68***	0.64
	(0.52)	(0.45)	(0.43)	
Real output (GDP)				
Nowcast	0.18	0.11	0.87***	0.69
	(0.22)	(0.22)	(0.18)	
1 quarter ahead	−0.07	0.54*	0.44*	0.35
	(0.39)	(0.29)	(0.23)	
2 quarters ahead	−0.20	0.29	0.67***	0.23
	(0.54)	(0.30)	(0.23)	
3 quarters ahead	1.01	−0.06	0.59**	0.08
	(0.81)	(0.37)	(0.25)	
4 quarters ahead	0.09	0.16	0.69***	0.11
	(0.83)	(0.31)	(0.23)	

Notes: Sample: 1968Q4 to 2006Q4 (about 150 observations were included).
* Indicate significance at the 10% level.
** Indicate significance at the 5% level.
*** Indicate significance at the 1% level.

Table 3
Timing test.

Horizon	δ	γ^P	γ^S	Adj. R^2
Inflation (GDP deflator)				
Nowcast	−0.29	0.99***	0.06	0.82
	(0.19)	(0.13)	(0.13)	
1 quarter ahead	−0.17	0.50***	0.55***	0.69
	(0.26)	(0.18)	(0.17)	
2 quarters ahead	−0.14	0.25	0.80***	0.67
	(0.28)	(0.15)	(0.13)	
3 quarters ahead	−0.10	0.11	0.93***	0.66
	(0.30)	(0.14)	(0.12)	
Real output (GDP)				
Nowcast	0.14	1.36***	−0.30**	0.66
	(0.25)	(0.11)	(0.12)	
1 quarter ahead	−0.21	1.04***	−0.01	0.33
	(0.41)	(0.22)	(0.22)	
2 quarters ahead	−0.38	0.67***	0.35*	0.21
	(0.54)	(0.24)	(0.18)	
3 quarters ahead	0.32	−0.06	0.87***	0.29
	(0.70)	(0.24)	(0.12)	

Notes: Sample: 1968Q4 to 2006Q4 (about 150 observations were included).
* Indicate significance at the 10% level.
** Indicate significance at the 5% level.
*** Indicate significance at the 1% level.

sample, the Greenbook forecasts possess additional information on inflation and output that is not contained in the SPF forecasts. All of the estimates of γ^S are significantly positive for all forecasting horizons considered, and the estimates of γ^P are mostly insignificant and close to zero. Only for the inflation nowcast and the one-period-ahead forecast of output do the SPF forecasts contain valuable information. Hence, including Greenbook forecasts would have improved the private forecasts.

When giving the Fed staff a timing disadvantage of one quarter and extending the forecast horizon in the test by one (see Eq. (5)), we find that the Greenbook forecasts no longer provide additional information for short-term forecasts, but are still useful for private forecasters at longer forecast horizons for both inflation and output. This might indicate that central banks actually do have a deeper understanding of the structural causes of inflation (and GDP growth), because this is required for the production of good forecasts at longer horizons. On the other hand, access to the most recent information is essential for a good nowcasting performance.

4. Analysing the dynamics of forecast performance

In this section, we investigate whether the relative performances of staff and private forecasts vary over time. Such could be the case as a result of changes in volatility patterns and given the presence of extraordinary uncertainty. Anecdotal evidence supports this point. First, Fed Chairman Bernanke mentioned in a press conference (on 12 December, 2012) that the Fed had overestimated real GDP growth in past years. Second, in the presence of persistent oil price shocks, several inflation-targeting central banks appear to have underestimated inflation for some time. The Stockton (2012) Report suggests that the Bank of England's forecast performance has deteriorated recently, being somewhat worse than that of private forecasters. Third, Kenny and Morgan (2011) document the predictive failure of forecasters' macroeconomic tools and expert judgement more broadly during the financial crisis for both short- and medium-term horizons. In this context, performance assessments by central banks indicate that wrong assumptions concerning oil prices and fiscal policies were at the root of the forecast errors.

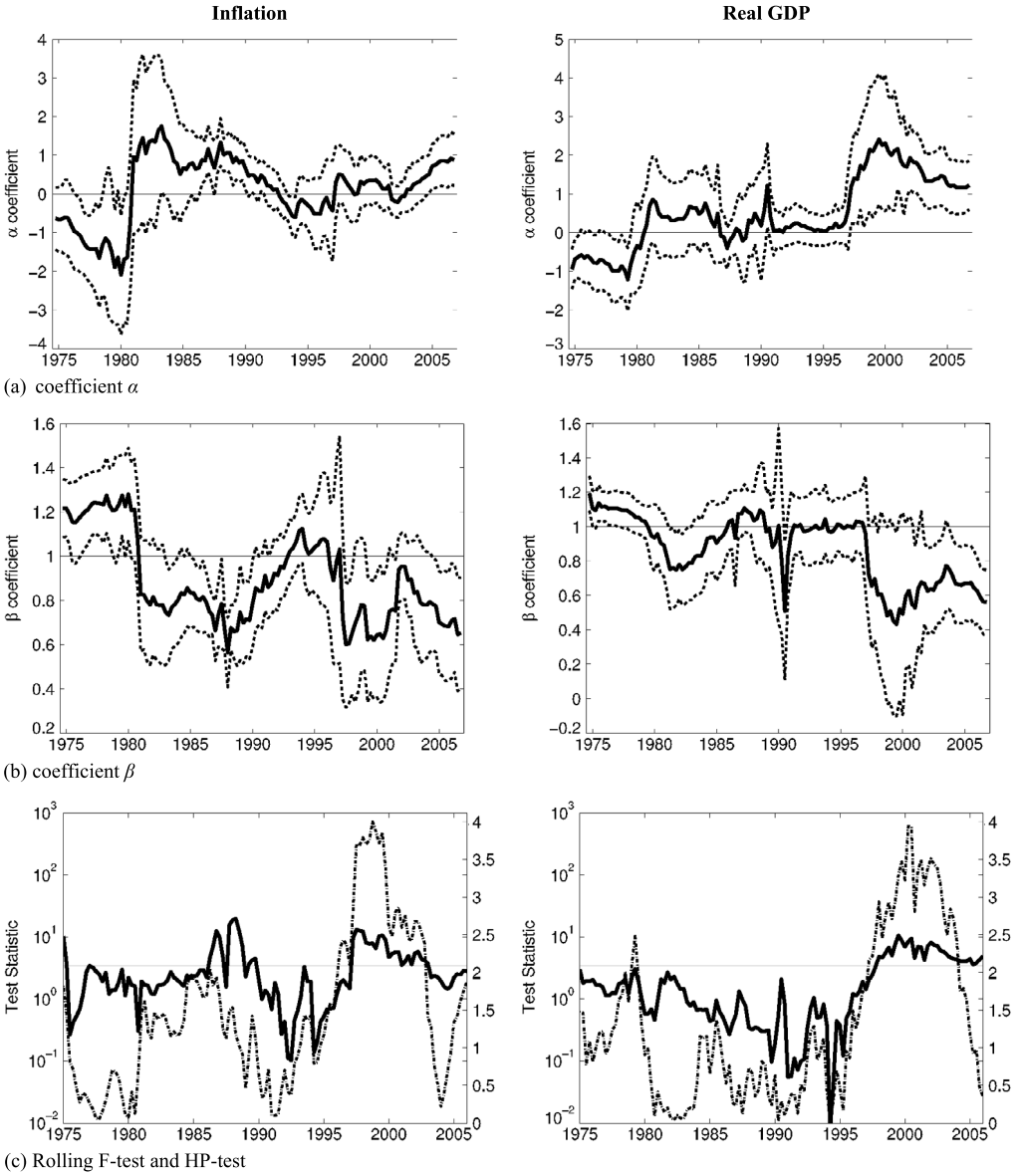
Section 4.1 examines time-varying efficiency and Section 4.2 presents tests of the stability of the relative forecast performances. Section 4.3 presents conditional predictive ability tests.

4.1. Time-varying efficiency (and unbiasedness)

4.1.1. Rolling window rationality

Rossi (2005) and Rossi and Sekhposyan (in press) argue that the tests described in Section 4 for forecast rationality are not valid in the presence of parameter instability.[4] Changes in US monetary policy paradigms may imply structural breaks in the relationship. In 1979, the Fed embarked on a disinflationary monetary policy. In an unusual announcement, chairman Volcker broke with past traditions and made it clear that the Fed would take responsibility for inflation (see Goodfriend, 1997, p. 12).

[4] This issue will be addressed in Section 4.2, where we apply fluctuation tests (see Giacomini & Rossi, 2010).



(a) coefficient α

(b) coefficient β

(c) Rolling F-test and HP-test

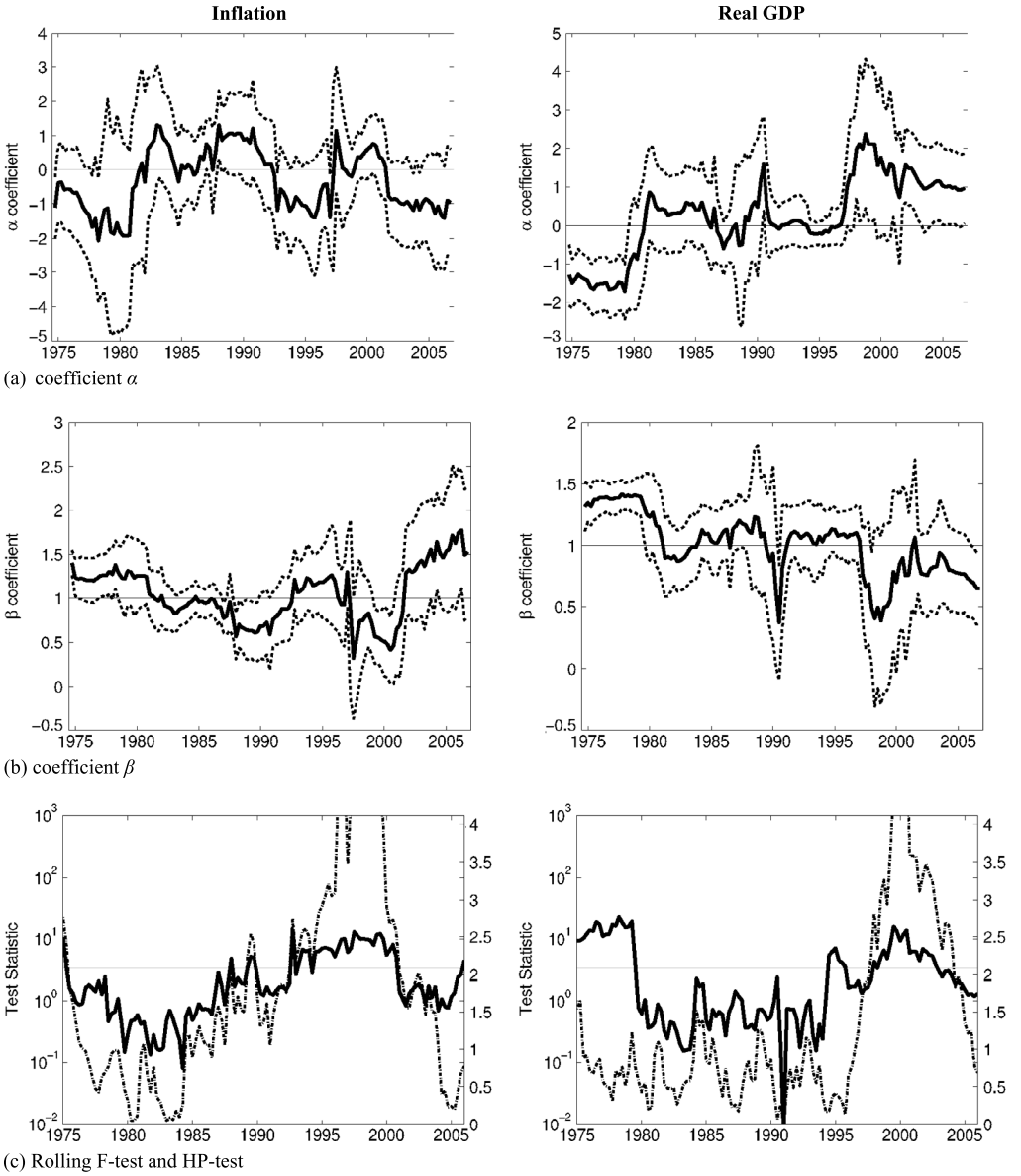
Note: The top four graphs show rolling window estimates for the individual α and β coefficients (solid lines), with corresponding 95% confidence bounds (dashed lines). The window size for the estimation is 25 observations. The two bottom graphs show the corresponding evolution of the F-statistic (for the joint hypothesis $\alpha = 0$ and $\beta = 1$; solid line) and the HP (Holden-Peel) test (dotted line). The axis is scaled so that the horizontal lines represent the critical values for both tests.

Fig. 1. Rolling window estimates: Fed Greenbook forecasts.

This was an important clarification, because it implied that the Fed would give more weight to price stability in the aftermath, within the dual mandate. The Volcker disinflation led to a regime shift towards lower inflation in the US. We show that this change has also had implications for the forecast rationality of both central bank staff and private forecasters. In order to check for the existence of a break in the relationship, we conduct a break point test,[5] which shows that a break probably occurred at the beginning of the 1980s, i.e. when the Volcker disinflation started (see the last column of Table 1). To account for those changes, we run our rationality tests for a moving window with a bandwidth of 25 quarters.

The rolling window estimates (see Figs. 1 and 2) show occasional or even prolonged departures from the unbiasedness property by both Greenbook and SPF forecasts. Efficiency is rejected for most periods and for all forecast horizons for both inflation and output. However, the above-mentioned phenomenon of the autoregressive

[5] We use the Andrews, Lee, and Ploberger (1996) procedure to test for breaks at an unknown time with a trimming parameter of 15%.



(a) coefficient α

(b) coefficient β

(c) Rolling F-test and HP-test

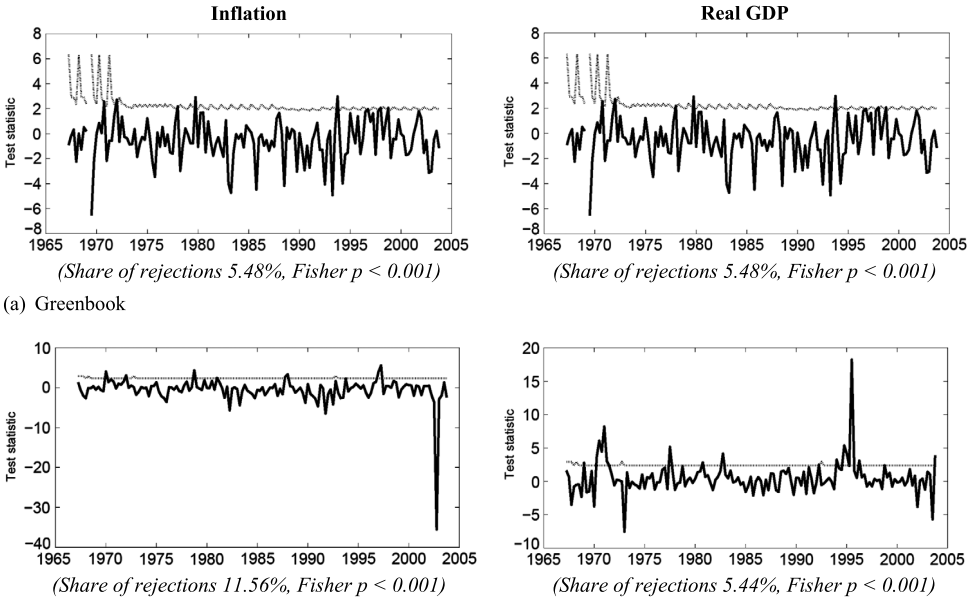
Note: The top four graphs show rolling window estimates for the individual α and β coefficients (solid lines), with corresponding 95% confidence bounds (dashed lines). The window size for the estimation is 25 observations. The two bottom graphs show the corresponding evolution of the F-statistic (for the joint hypothesis $\alpha = 0$ and $\beta = 1$; solid line) and the HP (Holden-Peel) test (dotted line). The axis is scaled so that the horizontal lines represent the critical values for both tests.

Fig. 2. Rolling window estimates: SPF forecasts.

behaviour of forecast errors obtained from rational models is particularly severe in small samples, as was shown by Evans and Lewis (1995), which would render the test invalid.

4.1.2. Fixed event efficiency test

Nordhaus (1987) argues that conventional tests of (weak) forecast efficiency ignore a crucial element of efficiency, namely the efficient updating of forecasts for a fixed event based on incoming information. The conventional tests for efficiency, as used in the previous sections, implicitly assess whether the forecast errors at time t have predictive power for the forecast errors at time $t + 1$ (after correcting for the correlation generated by overlapping forecast horizons). Hence, a persistence of expectations or forecasts is defined as repeating one's mistakes forecasts for future events. Nordhaus (1987) points out that it could be equally problematic if forecasters failed to include incoming information immediately when updating their forecasts, but were hesitant in revising



(a) Greenbook

(b) Survey of Professional Forecasters

Note: The graphs show the t-statistics of the AR coefficient of the Nordhaus test (which is a simple AR model of forecast revisions). The dashed line corresponds to the critical value for an individual test. Since the number of forecasts available for a given event fluctuates, the degrees of freedom vary between regressions, causing the critical value to fluctuate too. The share of rejections refers to a 5% significance level.

Fig. 3. Results of the Nordhaus test (t-statistics).

forecasts for a given event, so that new information seeped in only slowly. To assess this phenomenon empirically, he proposes the estimation of simple AR(1) models of forecast revisions for a fixed event. That is, we estimate an individual regression for every period for which forecasts are available in our sample. Since these regressions have very few degrees of freedom, and thus provide only limited information, we also aggregate the results using the Fisher test that has been reintroduced to modern econometrics by Maddala and Wu (1999) for the case of unit root tests. As with the rolling window rationality test, we find evidence of episodes of inefficiency (see Fig. 3). For both staff and private forecasters, the Fisher test rejects the joint null strongly for both inflation and GDP growth, indicating that the number and strength of the rejections are beyond what we might expect from a mere multiple testing problem.

4.2. Testing for forecast stability

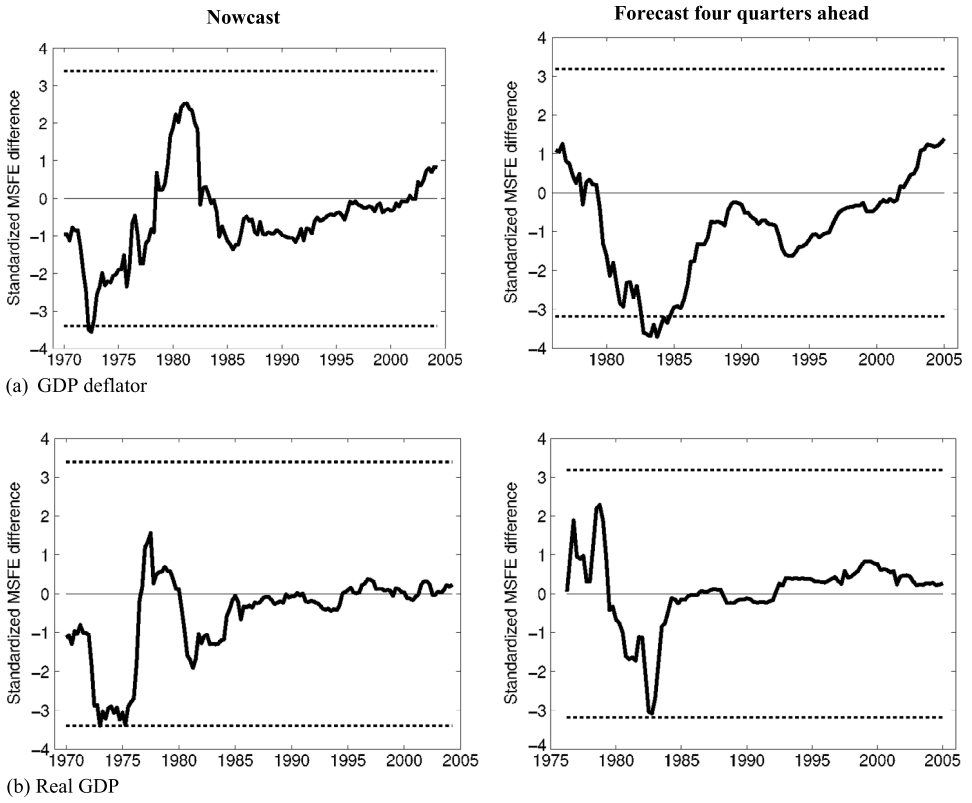
It is conceivable that there may be no differences in performances for certain episodes, but significant differences for other periods. We examine the stability of the relative forecast performances using the fluctuation test developed recently by Giacomini and Rossi (2010). The null hypothesis of the test is forecast stability, i.e. that the difference detected in the relative forecast performances is not time-varying:

$$H_0 : E[\Delta L_t(\hat{f}^S_{t-h,R}, \hat{f}^P_{t-h,R})] = 0, \quad \text{for all } t = R+h, \ldots, T, \tag{6}$$

where $\hat{f}_{t-h,R}$ denotes the h-step-ahead forecast errors at time t by Fed staff (superscript S) and the private forecasters (superscript P). L denotes the corresponding loss function. The test statistics are computed using rolling (out-of-sample) windows of a given size R. We choose the window size to equal 15% of the sample, as was suggested by Giacomini and Rossi (2010).

We perform the test for both inflation and GDP (also including the components of GDP). A graphical summary of the results for inflation (measured as the GDP deflator) and GDP growth is provided in Fig. 4. The null hypothesis of forecast stability is rejected if the test statistic hits the critical value at least once.[6] While we find that nearly all tests fail to reject the hypothesis of a stable relative forecast performance, the test rejects in several cases for the key variables of interest, i.e., inflation and growth, with two marginal rejections (i.e. for the nowcast for the GDP deflator and for the nowcast for real GDP) and one clear rejection (i.e. for the GDP deflator four quarters ahead). These breaks clearly coincide with beginning of the Great Moderation, which is not available in the samples for CPI inflation and the GDP components, and are probably attributable to the reduced volatility since the mid-1980s—i.e. after the US economy had faced severe oil price shocks in the 1970s.

[6] The supremum test suggested by Giacomini and Rossi (2010) already accounts for the multiple testing problem. The critical values for the individual Diebold–Mariano-type tests that are used to generate test statistics for individual points in time would be substantially lower.



Note: The solid line shows the fluctuation test statistic and the dashed lines represent the corresponding critical values. MSFE: mean squared forecast error.

Fig. 4. Fluctuation test statistic for inflation and output.

4.3. Testing for conditional predictive ability

The relative performances of different forecast groups may be explained by differences in the underlying assumptions. In this context, it has been argued that central bank staff have a more timely and complete knowledge of official statistics, and may be able to access data releases earlier than private forecasters. In addition, in the absence of forward guidance, central bank staff should have a better knowledge of the central bank's reaction to future shocks and the implied future interest rate path. The test by Giacomini and White (2006) allows us to account for the possibility that the forecast performance may be related to specific factors. We examine the predictive ability conditional on three key factors (data revisions, interest rate path, oil and commodity prices), and test for the influence of increased uncertainty, as measured by a volatility index.

The test provides information as to whether changes in the relative forecasting performance are linked to developments in specific exogenous variables. The null hypothesis is that, given the information set Ω_t, it is not possible to distinguish which forecast group has the lower forecast error at horizon τ. It can be written as:

$$H_0 : E[L(\hat{f}^S_{t+\tau}) - L(\hat{f}^P_{t+\tau})|\Omega_{t_t}] = 0. \quad (7)$$

First, we examine the relative forecasting performances for inflation and real GDP growth, given uncertainties in the economic environment, which are proxied by the cross-sectional dispersion of the quarterly forecasts (i.e. the dispersion of inflation, real GDP, industrial production, and housing starts).[7] Since the dispersion measures are obtained from the SPF, they could report an information disadvantage that is specific to professional forecasters, because the uncertainty perceived by individual forecasters cannot necessarily be deduced from their mutual disagreement. To check for the robustness of the results, we use the predicted variance of inflation obtained from a simple GARCH(1,1) model as an alternative proxy for economic uncertainties.

Second, we check whether data revisions had an impact on the relative forecasting performances, using revisions in the variables inflation and real GDP growth.[8] Third, to account for the Fed's better knowledge of its interest rate policy, we test for the impact of upcoming interest rate changes on relative forecast performances, using the absolute quarter-on-quarter change in the federal funds rate at the corresponding forecast horizon as a proxy.

[7] The dispersion measure equals the 75th percentile of the forecasts minus the 25th percentile, for quarter-on-quarter variables. These variables are available from the Federal Reserve Bank of Philadelphia.

[8] This set of revisions is obtained from the real-time data set of the Federal Reserve of Philadelphia.

Table 4
Tests for conditional predictive ability.

Horizon	Uncertainty					Data revisions		Interest rates	Commodity prices
	Inflation dispersion	Real output dispersion	Industrial production dispersion	Housing starts dispersion	ARCH measure	Inflation revisions	Real output revisions	Fed funds rate changes	HWWA-Index
Inflation (GDP deflator)									
Nowcast	0.359	0.226	0.264	0.478	0.270	0.400	0.254	0.112	0.919
1 quarter ahead	**0.025**	**0.022**	**0.002**	**0.049**	**0.027**	0.422	**0.025**	0.151	0.380
2 quarters ahead	**0.085**	**0.080**	**0.012**	0.193	0.135	0.814	0.148	0.108	0.367
3 quarters ahead	**0.052**	**0.018**	**0.016**	**0.000**	**0.001**	0.524	0.652	0.306	0.443
4 quarters ahead	**0.001**	**0.001**	**0.001**	**0.000**	**0.000**	0.556	0.178	**0.009**	0.494
Real output (GDP)									
Nowcast	**0.050**	**0.015**	**0.017**	**0.024**	**0.080**	**0.008**	0.295	**0.081**	0.307
1 quarter ahead	0.241	0.258	0.156	0.293	0.219	0.561	0.966	0.655	0.300
2 quarters ahead	0.878	0.757	0.950	0.820	0.850	0.231	0.595	0.808	0.304
3 quarters ahead	0.798	0.555	0.802	0.810	0.678	0.232	0.470	0.861	0.502
4 quarters ahead	0.767	0.906	0.728	0.519	0.822	0.357	0.342	0.975	0.287

Notes: *p*-values for the conditional predictive ability test are computed following Giacomini and White (2006). *p*-values below 10% are shown in bold. ARCH measures the volatility in the past inflation rate, and is used as an alternative uncertainty measure. Fed funds rate changes refers to quarter-on-quarter changes in the Fed's funds rate at the corresponding forecast horizon, and HWWA is an index comprising world market prices for energy, oil and raw materials.

Fourth, we condition on oil prices and the commodity price index.

Table 4 reports the results. First, if we condition the relative forecast performance on uncertainties in the economic environment, we find that Fed staff generally made better inflation forecasts than private forecasters (SPF) during times of increased economic uncertainty. The same holds for the nowcasts of output. However, this advantage cannot be detected for the nowcasts of inflation or for output forecasts for longer horizons.

Second, if we condition the relative forecast performances on data revisions, we find that the relative forecasting performances are affected significantly only in the very short term. Surprisingly, revisions in inflation cause improvements in the relative forecasting performances for real GDP, and revisions in real GDP cause improvements in the relative forecasting performances for inflation. Hence, for most horizons, the test results support Romer and Romer's (2000) argument that the Fed staff make better forecasts for reasons which are not related to their earlier access to government statistics.

Third, if we condition the relative forecast performances on future changes in the federal funds rate, the relative performances of Fed staff inflation forecasts are better at the longer horizon of four quarters ahead. This test suggests that the Fed probably made better inflation forecasts when interest rate changes were looming.[9] In that sense, the Fed's Greenbook inflation forecasts seem to have benefited from the staff's better knowledge of the Fed's future interest rate path.

Fourth, if we condition the relative forecast performances on oil prices and the HWWA index for energy, oil and raw materials, we find that these factors have no significant influence on the horserace between Fed staff and private forecasters.[10] These test results are consistent with common knowledge that the two types of forecasters face an even challenge when attempting to predict the consequences of changes in oil and commodity prices for inflation and output.

5. Conclusions

Several explanations for the Fed's information advantage have been proposed in the literature. There are three prominent explanations: (i) the Fed's thorough forecasting process, including a vast range of resources devoted to the forecasting of macroeconomic variables; (ii) the Fed's knowledge of its own probable policy actions and its comparative advantage in collecting detailed information about current and recent movements in the economy; and (iii) the Fed's privileged access to confidential data based on its bank supervisory authority. Of these, Romer and Romer (2000) reject the ideas of inside information by staff on the future interest rate path, an early access to government statistics, and a better knowledge of data revisions as possible explanations.

This paper examines whether any time variation in the (relative) forecast performances has occurred. It suggests that further qualifications of Romer and Romer's (2000) findings need to be made. On the one hand, this paper confirms their finding that the Fed has a significant information advantage for inflation and output forecasts over the

[9] Since interest rate changes may be more frequent in times of greater economic uncertainty, we also check whether there is any correlation between these variables that may drive our results, thus distorting the interpretation of our results. We find only a small correlation of the variables (0.30), implying that the results can be attributed mainly to the separate effect of the interest rate.

[10] This finding is robust to the use of different measures for oil. Since the HWWA index also comprises commodity prices, Table 4 only report the results for this measure.

extended sample 1968–2006. The result is robust to a possible timing advantage of the Fed's staff relative to private forecasters. Further evidence suggests the superiority of the Greenbook forecasts, particularly when the uncertainty is high. On the other hand, this paper differs from that of Romer and Romer (2000) when it comes to the assessment of the driving factors that explain the information advantage. First, this paper finds that the Fed staff's access to better information on the future fed funds rate explains a different relative performance for inflation forecasts. Second, the finding of the Fed's superiority is sensitive to the presence of large macroeconomic shocks such as the Great Moderation and oil price shocks. In this context, an interesting question is whether the non-availability of the Greenbook forecasts in real time ultimately explains their superiority over private forecasts. This question could be addressed in further research by comparing the evidence with that for other central banks, which publish their staff forecasts in real-time.

Acknowledgments

The authors thank Joseph Haubrich, Mikael Apel, Marianna Blix Grimaldi, Aaron Jackson, Axel Lindner, Claus Michelsen, Barbara Rossi, Oreste Tristani, and the participants of the Pacific Rim conference in Tokyo, the Verein für Socialpolitik, and the Annual Western Economic Association Conference in Seattle for useful comments and suggestions. The views expressed are those of the authors and do not necessarily reflect those of the Deutsche Bundesbank or the ECB. The authors remain responsible for any errors or omissions.

References

Andrews, D. W., Lee, I., & Ploberger, W. (1996). Optimal changepoint tests for normal linear regression. *Journal of Econometrics*, *70*(1), 9–38.

Bernanke, B. (2010). Monetary policy and the housing bubble. In *Address at the annual meeting of the American economic association. Atlanta, Georgia on 3 January 2010*.

Bhattacharjee, A., & Gelain, P. (2011). Monetary policy under model uncertainty: a Bayesian analysis of FOMC forecasts. University of Dundee working paper.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (2008). Information cascades. In S. N. Durlauf, & L. E. Blume (Eds.), *The new Palgrave dictionary of economics* (2nd ed.). UK: Palgrave Macmillan.

Blattner, T., Catenaro, M., Ehrmann, M., Strauch, R., & Turunen, J. (2008). The predictability of monetary policy. ECB occasional paper series 83.

Brand, C., Buncic, D., & Turunen, J. (2010). The impact of ECB monetary policy decisions and communication on the yield curve. *J. Eur. Econ. Assoc.*, *8*(6), 1266–1298.

D'Agostino, A., & Whelan, K. (2008). Federal Reserve information during the great moderation. *J. Eur. Econ. Assoc.*, *6*(2–3), 609–620.

Ellison, M., & Sargent, T. J. (2012). A defense of the FOMC. *International Economic Review*, *53*(4), 1047–1065.

Evans, M. D., & Lewis, K. K. (1995). Do expected shifts in inflation affect estimates of the long-run Fisher relation? *Journal of Finance*, *50*, 225–253.

Gamber, E. N., & Smith, J. K. (2009). Are the Fed's inflation forecasts still superior to the private sector's? *Journal of Macroeconomics*, *31*(2), 240–251.

Gavin, W. T., & Mandal, R. J. (2003). Evaluating FOMC forecasts. *International Journal of Forecasting*, *19*(4), 655–667.

Giacomini, R., & Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, *25*(4), 595–620.

Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, *74*(6), 1545–1578.

Goodfriend, M. (1997). Monetary policy comes of age: a 20th century odyssey. *Federal Reserve Bank of Richmond Economic Quarterly*, *83*(1), 1–22.

Holden, K., & Peel, D. A. (1990). On testing for unbiasedness and efficiency of forecasts. *The Manchester School*, *58*(2), 120–127.

Jung, A. (2013). Policy-makers' interest rate preferences: recent evidence for three monetary policy committees. *International Journal of Central Banking*, *9*(3), 145–192.

Keane, M. P., & Runkle, D. E. (1990). Testing the rationality of price forecasts: new evidence from panel data. *American Economic Review*, *80*(4), 714–735.

Kenny, G., & Morgan, J. (2011). Some lessons from the financial crisis for the economic analysis. ECB occasional paper series 130.

Lange, J., Sack, B. P., & Whitesell, W. C. (2003). Anticipations of monetary policy in financial markets. *Journal of Money, Credit and Banking*, *35*(6), 889–909.

Maddala, G., & Wu, S. (1999). A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and Statistics*, *61*, 631–652.

McCracken, M. W. (2010). Disagreement at the FOMC: the dissenting votes are just part of the story. *The Regional Economist*, 10–16. October.

Nordhaus, W. (1987). Forecasting efficiency: concepts and applications. *Review of Economics and Statistics*, *69*, 667–674.

Nunes, R. (2013). Do central banks' forecasts take into account public opinion and views? International finance discussion papers no. 2013-1080.

Peek, J., Rosengren, E. S., & Tootell, G. M. B. (2003). Does the Federal Reserve possess an exploitable informational advantage? *Journal of Monetary Economics*, *50*(4), 817–839.

Reifschneider, D., & Tulip, P. (2007). Gauging the uncertainty of the economic outlook from historical forecasting errors. FEDS papers 2207-60, Board of Governors of the Federal Reserve System.

Romer, C. D., & Romer, D. H. (2000). Federal Reserve information and the behavior of interest rates. *American Economic Review*, *90*(3), 429–457.

Romer, C. D., & Romer, D. H. (2008). The FOMC versus the staff: where can monetary policymakers add value? *American Economic Review: Papers and Proceedings*, *98*(2), 230–235.

Rossi, B. (2005). Optimal tests for nested model selection with underlying parameter instabilities. *Econometric Theory*, *21*(5), 962–990.

Rossi, B., & Sekhposyan, T. (2015). Forecast rationality tests in the presence of instabilities, with applications to Federal Reserve and survey forecasts. *Journal of Applied Econometrics*, (in press).

Sheng, X. (2015). Evaluating the economic forecasts of FOMC members. *International Journal of Forecasting*, *31*, 165–175.

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

Stockton, D. (2012). *Review of the Monetary Policy Committee's forecasting capability. Technical report*. Bank of England.

Swanson, E. T. (2006). Have increases in federal reserve transparency improved private sector interest rate forecasts? *Journal of Money, Credit and Banking*, *38*(3), 791–819.

Tillmann, P. (2011). Strategic forecasting on the FOMC. *European Journal of Political Economy*, *27*(3), 547–553.

Tulip, P. (2009). Has the economy become more predictable? Changes in Greenbook forecast accuracy. *Journal of Money, Credit and Banking*, *41*(6), 1217–1231.

Woodford, M. (2005). Central bank communication and policy effectiveness. In *Proceedings of the Federal Reserve Bank of Kansas city, August* (pp. 399-474).

Makram El-Shagi is visiting Associate Professor at the California State University Long Beach, senior economist at the Halle Institute for Economic Research and Adjunct Professor at the University of Mannheim. He holds a Ph.D. as well as a habilitation (the German postdoctoral degree qualifying for full professorship) from the University of Mannheim.

Sebastian Giesen holds a Ph.D. from the Martin-Luther-University Halle-Wittenberg. Currently, he is economist in the Monetary Policy and Analysis Division of the Deutsche Bundesbank.

Alexander Jung is Principal Economist in the Directorate Monetary Policy of the European Central Bank. He has been working for more than 20 years in central banking and obtained his Ph.D. from the University of St. Gallen (Switzerland).