

Does Central Bank Staff Beat Private Forecasters?*

Makram El-Shagi[†] Sebastian Giesen[‡] Alexander Jung[§]

June 22, 2012

Abstract

In the tradition of Romer and Romer (2000) this paper compares staff forecasts of the Federal Reserve and the European Central Bank for inflation and output with corresponding private forecasts. Standard tests show that the FOMC and less so the ECB have a considerable information advantage about inflation and output. Novel tests for conditional predictive ability and forecast stability for the US (see Giacomini and White (2006) and Giacomini and Rossi (2010)), indicate that gradually, coinciding with the Great Moderation, the information advantage of Greenbook forecasts narrowed considerably.

Keywords: Relative forecast performance, forecast stability, staff forecasts, private forecasts, real-time data

JEL Classification: C53, E37, E52, E58

*The authors thank Mikael Apel, Marianna Blix Grimaldi, Axel Lindner, Claus Michelsen, Barbara Rossi, and the participants of the ECB forum on central bank decision-making and communication for useful comments and suggestions. The views expressed by Jung are his own and do not necessarily reflect those of the ECB. The authors remain responsible for any errors or omissions.

[†]Halle Institute for Economic Research, Makram.El-Shagi@iwh-halle.de.

[‡]Halle Institute for Economic Research, Sebastian.Giesen@iwh-halle.de.

[§]European Central Bank, Alexander.Jung@ecb.int.

1 Introduction

Forecast performance is a subject of more general interest in science. As the Danish physicist Niels Bohr famously said "prediction is very difficult, especially if it's about the future." Most economists have been concerned with forecast rationality and have asked the question whether forecasters correctly use any relevant information they knew in making their predictions. In other words are forecast errors unpredictable?

Private forecasters could improve their own inflation forecasts by gaining information about the Fed's Greenbook forecasts and by putting a large weight on them. This is the finding of a widely quoted study by Romer and Romer (2000) who argue that the Greenbook forecasts provide the FOMC with an information advantage relative to private forecasters. Learning about the Fed's forecasts in real-time, however, is hindered by the fact that these forecasts become publicly available with a lag of about 5 years. Strikingly, the published FOMC forecasts offer no alternative for private forecasters, because the information advantage does not apply to them (see Romer and Romer (2008)).

By now, it is widely acknowledged that for quite sometime central bank staff made better macroeconomic forecasts than private forecasters. This was owing to several factors including expert knowledge and a leading edge in modeling economies. After the onset of the Great Moderation in the mid-1980s, changes in the forecast ability of inflation and output in the United States and elsewhere have been observed. On the one hand, Stock and Watson (2007) argue that over time it has become more difficult to forecast inflation

in the US even though, apart from the crisis, the inflation process has become less volatile. On the other hand, Tulip (2009) finds that inflation and output have become more predictable, even though the results for output are somewhat mixed.

The aim of this paper is to compare staff forecasts of the Federal Reserve (Fed) and the European Central Bank (ECB) for inflation and output with corresponding private forecasts. Standard tests show that the FOMC and less so the ECB have a considerable information advantage about inflation and output. Novel tests for conditional predictive ability and forecast stability for the US (by Giacomini and White (2006) and Giacomini and Rossi (2010)) indicate that gradually, coinciding with the Great Moderation, the information advantage of Greenbook forecasts has narrowed considerably.

Section 2 briefly reviews the literature. Section 3 explains the data used in this paper. Section 4 analyzes the forecasting performance of the Fed's and ECB's staff forecasts relative to those by private forecasters. Section 5 concludes.

2 A brief review of the literature

Today, there is broad agreement on the principles of sound monetary policy. For example, to be effective monetary policy has to be forward-looking and has to take into account a wide range of indicators. In terms of transparency about the monetary policy framework and process, central banks across the world have made different choices. For many central banks inflation and

output forecasts are important elements for assessing the monetary policy stance in real time and explaining monetary policy to the public. Since the mid-1990s, most inflation targeting central banks have regularly provided the public with their staff forecasts when explaining their monetary policy decisions. The Fed releases its Greenbook staff forecasts only with a lag of about five years and communicates FOMC policy-makers' forecasts four times a year. Since July 1979, the Fed regularly publishes summary statistics of FOMC policy-makers' economic projections twice a year (in February and July) and since October 2007 four times a year (this is done in connection with the FOMC's policy meetings in January, April, June, and November). In a study, Romer and Romer (2008) compare the forecasts of the FOMC members with those from the staff. They find that these forecasts do not provide useful information relative to the Greenbook forecast even though FOMC members know the staff forecast when making their individual forecast. Ellison and Sargent (2009) argue that it would be difficult to justify inferior FOMC forecasts, when evidence suggests that differences between them and Greenbook forecasts are very small. Correctly interpreted, the FOMC forecast has a larger dispersion around the mean, because individual policy-makers respond differently to multiple uncertainties than Fed staff. These forecasts are possibly also influenced by other factors, such as strategic motives in forecasting (see McCracken (2010) and Tillmann (2011)). Romer and Romer (2000) find that the internal Greenbook forecasts provide the FOMC with an information advantage relative to private forecasters. They argue that the thorough forecasting process including a vast range of resources from the Fed staff is the explanation for the difference in behavior.

Hence, Greenbook forecasts provide the FOMC with an information advantage relative to private forecasters in the following sense. First, Greenbook forecasts have lower root mean square errors (RMSE) than private forecasts. Second, given the Fed's Greenbook forecast, private sector forecasts have little or no additional explanatory power for inflation.

Why should staff forecasts be superior to private forecasts? In fact, this is a puzzling result, because the level of data and model uncertainty is profound for both central bank staff and private forecasters. Sims (2002) suggests the Fed's forecasting advantage is attributable to the Fed's knowledge of its own likely policy actions and the Fed's comparative advantage in collecting detailed information about current and recent movements in the economy. A study by Peek, Rosengren and Tootell (2003) explains the finding by the Fed's privileged access to confidential data based on its bank supervisory authority. By contrast, Romer and Romer (2000) reject inside information by staff on the future interest rate path, the early access to government statistics and the better knowledge about data revisions as possible explanations. The rejection of the knowledge about the future path is consistent with the observation that in the past Greenbook forecasts were based on appropriate monetary policy, but in practice they often used the constant interest rate assumption subject to judgemental adjustment.

For reasons related to data availability, most studies examining this issue have been made for the US. Table 1 provides an overview of existing studies on the relative forecasting performance. For different samples ranging from the late 1960s to the mid-1990s several studies support the finding on the information advantage of the Fed (see Sims (2002), Gavin and Mandal (2003),

Table 1: Findings from the literature

Author of the study	Method	Sample	Finding
United States			
Romer and Romer (2000)	Rationality test, Encompassing test, MSEs	1968 - 1991	Greenbook forecasts for inflation and output are superior to private forecasts
Sims (2002)	RMSE, Factor analysis	1979 - 1995	Greenbook forecasts for inflation are superior; for output no significant advantage
Gavin and Mandal (2003)	RMSE	1979 - 1996	Greenbook forecasts for inflation are more accurate than private forecasts; the finding does not apply to output
Peek, Rosengren and Tootell (2003)	Encompassing test	1977 - 1996	Greenbook forecasts for inflation and output are superior to private forecasts
Reifschneider and Tulip (2007)	RMSFE, Test for predictive accuracy	1986 - 2006	Greenbook forecasts and private forecasts have broadly similar accuracy
D'Agostino and Whelan (2008)	MSFE, Encompassing test	1974 - 1991	Greenbook forecasts are only superior for inflation during 1974 to 1991; thereafter information advantage is reduced
Gamber and Smith (2009)	RMSE; Encompassing test	1968 - 2001	Information advantage of Greenbook forecasts is reduced since mid-1980s, especially after 1994
Rossi and Sekhposyan (2011)	MSFE, Rationality test, Encompassing test, Fluctuation rationality test	1968 - 2005	Information advantage of Greenbook inflation forecasts deteriorates after 2003
Euro area			
Genre, Kenny, Meyler and Timmermann (2010)	MFE, RMSFE	1999 - 2009	Pooling private forecasts can improve inflation forecasts, but not output forecasts
Hubert (2012)	RMSFE	2004 - 2011	ECB inflation forecasts outperform private forecasts for current year; for the next year no information advantage

Notes: MSE: mean square error, RMSE: root mean square error, MFE: mean forecast error, RMSFE: root mean square forecast error.

Peek et al. (2003)). The study by D'Agostino and Whelan (2008) finds that the information advantage of the Fed only applies to inflation forecasts and not to output forecasts. By contrast, Reifschneider and Tulip (2007) find that Greenbook forecasts since 1986 are not more accurate than private forecasts. Other recent studies for an extended sample up to 2006 by Gamber and Smith (2009) and Rossi and Sekhpoysan (2011) document a narrowing of the gap in forecast performance and hence a reduced information advantage of the Fed. This trend may be explained by several developments. Most notably, improved transparency standards on monetary policy implied that policy signals were better understood and anticipated by markets. Private forecasters have discovered the importance of forecasts for understanding monetary policy. Since the mid-1980s, private forecasters have continuously improved their forecasting methods and many of them apply state of the art forecasting models. Private research institutes also spend vast resources to make high quality forecasts and often publish these forecasts. In addition, the presence of information cascades has contributed to a better sharing of information and to a reduction in the dispersion of private forecasts.

Surprisingly, for the euro area the empirical literature on forecast comparison is scant. A recent study by Hubert (2012) finds that the ECB's inflation forecasts outperform private forecasts for the current year. Though, concerning forecasts for the next year this information advantage of the ECB staff forecast cannot be detected. Important differences between the ECB's staff forecast and the Fed's Greenbook forecast could explain the different behavior for longer forecast horizons.¹ First, the ECB regularly publishes its staff

¹There are also important similarities between both central banks. First, the forecast-

forecast, whereas the Greenbook forecast is not available to private forecasters in real-time. Second, the Fed staff conditions its forecast on appropriate monetary policy, and the ECB initially used the constant interest rate assumption and since June 2006 switched to a market interest rate assumption. Furthermore, during the financial crisis, owing to the extraordinary volatility and uncertainty, the forecast performance of most other central banks, and of private forecasters has significantly deteriorated (see Kenny and Morgan (2011)). As suggested by Genre, Kenny, Meyler and Timmermann (2010) pooling private forecasts from the SPF may help in these circumstances, and lead to better outcomes for forecasts on inflation.

3 A real-time database

In order to perform econometric tests on forecast efficiency and on their predictive ability, it is necessary to collect real-time data on forecasts and their outcomes. This paper uses quarterly data for the US from 1968Q4 to 2006Q4 and for the euro area from 2000Q4 to 2012Q1. We focus on inflation and output and use the Survey of Professional Forecasters (SPF) to proxy private forecasters. In the following we document data used separately for the United States and for the euro area.

Concerning the US, we use data from the Greenbook and the SPF, as well as real-time data on eight key economic variables from the real-time database

ing process at both central banks involves a wide range of staff and several econometric models. Though, at the ECB the number of experts involved is higher, because also experts from the National Central Banks (NCBs) are involved in it (see Moutot, Jung and Mongelli (2008)). Second, like other central banks, the ECB internally tracks its forecast performance relative to other forecasters with the aim to improve forecasting accuracy and methods.

of the Federal Reserve Bank of Philadelphia. These include real GDP and its components (i.e. real consumption, real fixed business investment, real residential investment, real federal government consumption, real local and state government consumption), nominal GDP and the GDP deflator. The Greenbook projections are prepared by the research staff at the Board of Governors and are produced before each meeting of the FOMC. The SPF is the oldest quarterly survey of macroeconomic forecasts in the US and has been conducted by the American Statistical Association and the National Bureau of Economic Research. The Bank of Philadelphia took over the survey in 1990. When comparing the forecasts from the Greenbook with the SPF forecasts it should be taken into account that the SPF forecasters have a slight timing advantage over the Federal Reserve Board staff who prepare the Greenbook projections.

Concerning the euro area, we use publicly available historical series of the Eurosystem's staff macroeconomic projections for inflation and output in real time as well as the corresponding projections from the SPF (both from the ECB website). The corresponding real-time database on economic variables is taken from the statistical data warehouse (see Giannone, Henry, Lalik and Modugno (2010) for a detailed description). Projections in June and December are prepared by the Eurosystem staff and the projections in March and September are prepared by the ECB staff. Projections are published for the current year, the next year, and (in December) two years ahead. However, in contrast to the US data set, disaggregated data for the SPF is not available.

4 Econometric strategy and results

In this section we analyze the forecasting performance of central banks' staff relative to those by private forecasters. First, we test for rationality of the various forecasters based on the full and on sub-sample regressions. Second, we test whether central banks staff possess information that is potentially helpful to improve private sector forecasts. Third, we test whether or not the relative performance of central bank and private forecasts is stable over time. Fourth, we identify the reasons for the detected fluctuations in the relative forecasting performance by using alternative exogenous variables.

4.1 Forecast rationality

Tests on forecast rationality are commonly used by researchers to check whether forecasters make use of available information in a reasonably efficient way. To address this issue we use standard tests (see Romer and Romer (2000)) and regress the h-period ahead outcome on the h-period ahead forecasts at time t for inflation and real GDP:

$$A_{h,t} = \alpha + \beta * \hat{F}_{h,t} + \varepsilon_{h,t}, \quad (1)$$

where $A_{h,t}$ denotes outcomes (inflation rate or real GDP growth rate) h-steps ahead, and $\hat{F}_{h,t}$ is the corresponding h-step ahead forecast (of the inflation rate or the real GDP growth rate). If α equals zero and β equals one the forecast is called rational. The null hypothesis of rationality is $H_0 : \alpha =$

0 and $\beta = 1$ jointly. To deal with the potential problem of serial correlation in the forecast errors when estimating equation 1, we calculate robust standard errors for all regressions.

Our results for the US confirm those obtained by Romer and Romer (2000) for an extended sample (see Table 2). We find that the null hypothesis of forecast rationality cannot be rejected at conventional levels for both the Greenbook forecasts and the SPF. The only noteworthy exception are the growth forecasts at three or four quarters ahead, as indicated by the low p-values and the deterioration in R^2 . At this horizon the β coefficient is smaller than unity, whereas for inflation forecasts the β coefficient is still close to unity. Overall, the tests confirm that Greenbook and private forecasts contain important information about future inflation and output. This finding is in line with Rossi and Sekhpoysan (2011) who find evidence for rationality for both forecasters, including the private Blue Chip forecasts. Due to the fact that there is no Blue Chip data available for the euro area we stick to the SPF in the present comparison.

Table 2: Forecast rationality - United States

Horizon (quarters)	α	β	p-value	R^2	N	BP-test
Inflation						
Fed						
0	0.07(0.22)	0.96(0.06)	0.85	0.67	152	80Q4
1	0.06(0.29)	1.00(0.08)	0.90	0.75	152	80Q4
2	0.08(0.30)	1.01(0.09)	0.80	0.68	150	80Q3
3	0.04(0.33)	1.01(0.10)	0.92	0.65	144	80Q2
4	0.01(0.38)	1.01(0.12)	0.99	0.61	137	80Q1
SPF						
0	-0.30(0.21)	1.05(0.06)	0.30	0.82	152	81Q1
1	-0.21(0.31)	1.05(0.09)	0.78	0.68	152	80Q4
2	-0.20(0.36)	1.04(0.10)	0.86	0.59	152	80Q4
3	-0.16(0.47)	1.03(0.13)	0.94	0.51	152	81Q2
4	-0.02(0.55)	0.99(0.14)	0.98	0.43	147	81Q1
GDP						
Fed						
0	0.24(0.20)	0.96(0.05)	0.49	0.64	152	79Q3
1	0.30(0.34)	0.83(0.10)	0.16	0.33	152	79Q2
2	0.10(0.44)	0.85(0.13)	0.20	0.19	150	80Q1
3	0.92(0.54)	0.56(0.16)	0.01	0.06	144	79Q4
4	0.83(0.61)	0.63(0.19)	0.09	0.06	137	92Q4
SPF						
0	-0.11(0.23)	1.12(0.07)	0.10	0.64	152	79Q3
1	-0.21(0.39)	1.03(0.12)	0.83	0.33	152	79Q2
2	-0.32(0.39)	1.00(0.17)	0.37	0.19	152	79Q1
3	0.33(0.81)	0.72(0.22)	0.07	0.06	152	79Q4
4	0.05(0.87)	0.79(0.26)	0.04	0.06	152	82Q1

Notes: The table comprises test statistics for the rationality of the inflation forecasts. The standard errors are computed using the Newey-West procedure. BP-test denotes the detected break point date using a standard Andrews-Lee-Ploberger breakpoint test.

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)). This was an important clarification, because it implied that in the aftermath the Fed would give more weight to price stability within the dual mandate and at the same time emphasized the key role of monetary aggregates as driver for inflation. In order to check for the existence of a break in the relationship we conduct a break point test.² A break has likely occurred at the beginning of the 1980s, i.e. when the Volcker disinflation started (see last column of Table 2).

Rossi (2005) and Rossi and Sekhpoysan (2011) argue that the above standard tests for forecast rationality are invalid in the presence of parameter instability.³ Changes in paradigms of US monetary policy may imply structural breaks in the relationship. The Volcker disinflation has 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. Other factors such as the Asian-Russian crisis in 1998 may also have contributed to the violation of forecast rationality. Using rolling window estimation techniques Figure 2 and 3 indicate occasional or even prolonged departures from the rationality property, which are more pronounced for inflation forecasts than for output forecasts.

Table 3 shows the results for the euro area. Like for the US the tests tend to confirm forecast rationality both for the ECB staff and for private forecasters, although the results for inflation are rather weak in comparison to those for the US. With one exception, the null of forecast rationality cannot be rejected at the five percent level using a Wald test. Given that the regression may suffer from a

²We used the Andrews, Lee and Ploberger (1996) procedure to test for breaks at unknown time with a trimming parameter of 15%.

³We address this issue also in section 4.3 by applying fluctuation tests (see Giacomini and Rossi (2010)).

Table 3: Forecast rationality - Euro area

Horizon (years)	α	β	p-value	R^2	N
Inflation					
ECB					
0	0.07(0.05)	0.98(0.03)	0.08	0.96	38
1	0.14(0.83)	1.07(0.46)	0.08	0.64	36
SPF					
0	0.16(0.14)	0.98(0.07)	0.04	0.89	39
1	-0.15(1.02)	1.21(0.54)	0.11	0.38	36
GDP					
ECB					
0	-0.06(0.07)	1.03(0.04)	0.71	0.95	37
1	-0.88(0.57)	1.10(0.29)	0.11	0.43	34
SPF					
0	-0.26(0.10)	1.10(0.07)	0.05	0.93	37
1	-0.69(0.77)	0.96(0.37)	0.15	0.35	34

Notes: The table comprises test statistics for the rationality of the euro area forecasts. The standard errors are computed using the Newey-West procedure. Estimates for the euro area are obtained using DOLS.

small sample bias ($n \leq 40$), we apply a dynamic ordinary least squares (DOLS) estimator technique to estimate the parameters.⁴ Moreover, the ECB changed its interest rate assumption used to condition its staff forecasts in June 2006 from constant rates to market expectations. This change could imply a structural break for the staff forecast. Using the rolling window forecasting technique Figure 4 and 5 show temporary violations of the forecast rationality property in the case of the euro area for both ECB staff and private forecasters. Hence, the change in the

⁴See Stock and Watson (1993) who show that DOLS corrects for small sample bias and leads to unbiased and efficient parameter estimates.

forecast assumption was not responsible for the break. Like for the US violations are more persistent for inflation than for output forecasts. Interestingly, during the period of the financial crisis euro area inflation forecasts the test properties for rationality improved. Only for the ECB staff forecasts this development is strong enough that rationality is no longer rejected at the 5% significance level. For the private forecasters rationality is still rejected at this level. Two explanations for this observation are conceivable. First, noise during the crisis lead to the increased confidence bounds of the point estimates of α and β , making it harder for the rationality test to reject. Second, the point estimates move towards $\alpha = 0$ and $\beta = 1$, i.e. the parameters implying rationality. The reason for the changing parameter estimates is, that the shocks also caused strong deviation of macroeconomic indicators from their long run mean. These deviations could in turn be exploited as information for the subsequent forecasts. Visual inspection of the test results implies that the latter effect dominated the change in results at least for the ECB. The rationality of forecasts during the crisis does not imply that forecast errors decreased during the crisis (see Kenny and Morgan (2011)).

4.2 Testing for additional information

In this section we test whether the central bank staff forecasts actually possess additional information about the current and future economic environment which private forecasters could use to improve their forecasts (encompassing test). Central banks could have information about the economy that is not known to market participants when preparing their inflation and output forecasts. To test for such additional information, as proposed by Romer and Romer (2000), we estimate the following equation:

Table 4: Encompassing test - United States

Horizon (quarters)	δ	γ^P	γ^S	R^2	N
Inflation					
0	-0.11(0.19)	0.32(0.11) ***	0.69(0.13) ***	0.85	152
1	0.11(0.33)	-0.09(0.17)	1.08(0.16) ***	0.75	152
2	0.26(0.36)	-0.26(0.35)	1.23(0.36) ***	0.68	150
3	0.21(0.47)	-0.21(0.35)	1.19(0.32) **	0.65	144
4	0.58(0.52)	-0.77(0.45) *	1.68(0.43) ***	0.64	136
GDP					
0	0.18(0.22)	0.11(0.22)	0.87(0.18) ***	0.69	152
1	-0.07(0.39)	0.54(0.29) *	0.44(0.23) *	0.35	152
2	-0.20(0.54)	0.29(0.30)	0.67(0.23) ***	0.23	150
3	1.01(0.81)	-0.06(0.37)	0.59(0.25) **	0.08	144
4	0.09(0.83)	0.16(0.31)	0.69(0.23) ***	0.11	136

Notes: The table comprises coefficient estimates and corresponding test statistics for the test of staffs additional information for inflation. The asterisk marks significance at the one(***) , five(**) and ten(*) percent level.

$$A_{h,t} = \delta + \gamma^P * \hat{F}_{h,t}^P + \gamma^S * \hat{F}_{h,t}^S + \nu_{h,t}, \quad (2)$$

where $A_{h,t}$ denotes outcomes (inflation rate or real GDP growth) h-steps ahead, and $\hat{F}_{h,t}$ is the corresponding h-step ahead forecast from the central bank staff (superscript S) or the private forecaster (superscript P).

The existence of additional information by central bank staff would require that γ^S is positive and significantly different from zero. Table 4 reports estimation results of equation 2 for the US. Like Romer and Romer (2000), our results indicate that

Table 5: Encompassing test - Euro area

Horizon (years)	δ	γ^P	γ^S	R^2	N
Inflation					
0	0.23(0.10)	-0.01(0.23)	0.92(0.21) ***	0.93	39
1	2.78(1.53) *	-0.73(1.21)	0.32(0.54)	0.04	36
GDP					
0	-0.02(0.08)	0.04(0.25)	0.93(0.24) ***	0.94	39
1	0.48(1.47)	-2.57(2.29)	2.94(1.88)	0.23	36

Notes: The table comprises coefficient estimates and corresponding test statistics for the test of staffs additional information for inflation. The estimations are carried out using DOLS. The lead and lag length is thereby determined using the Akaike information criterion. The asterisk marks significance at the one(***), five(**) and ten(*) percent level.

for an extended sample Greenbook forecasts possess additional information on inflation and output which is not contained in the SPF forecasts. All 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 nowcast on inflation and the one-period-ahead forecast of output the SPF forecast contains valuable information. Hence, including Greenbook forecasts would have improved private forecasts.

Table 5 reports estimation results for the euro area which are more mixed. For the current year the tests clearly indicate that the ECB also possesses additional information which is not contained in the SPF forecasts. However, there seems to be no such additional information for next years forecasts. An issue here are the low values of R^2 for the year ahead forecast. This deterioration could be a

reflection of the extraordinary uncertainties with which forecasters had to cope during the financial crises since 2007. It could also be due to the fact that market participants have the opportunity to include the information on the current economic environment incorporated in staff projections in case of the ECB but not for the Fed.

4.3 Testing for forecast stability

In this section, we examine forecasting performance and stability of Greenbook forecasts relative to private forecasts over the last decades using a novel test by Giacomini and Rossi (2010). The test can only be applied for the US in a meaningful way because it requires sufficiently long datasets. Its null hypothesis is forecast stability:

$$H_0 : E[\Delta L_t(\hat{f}_{t-h,R}^S, \hat{f}_{t-h,R}^P)] = 0 \quad \text{for all} \quad t = R + h, \dots, T, \quad (3)$$

where $\hat{f}_{t-h,R}$ denotes the h -step ahead forecast errors from the central bank staff (superscript S) and the private forecasters (superscript P). L denotes the corresponding loss function. The test statistics are computed by means of rolling out-of-sample windows of a given size.⁵

The null hypothesis of forecast stability is rejected, if the test statistics hits one of the confidence bounds shown in Figure 1. This is the case for inflation as well as for real GDP growth but not for the GDP components. When analyzing the GDP components it turns out that some marked fluctuations for private consumption and investment are observable. Here, the test indicates no instability in the re-

⁵For details see Giacomini and Rossi (2010) equation (1). As they suggest, we choose the window size to equal 15% of the sample.

relationship and that in most instances the information advantage is on the side of the Fed forecasts.

Figure 1 shows the results for inflation, output and its components for the US. Negative (positive) values indicate the superiority of Fed Greenbook (SPF) forecasts. The Greenbook nowcasts significantly outperformed the private forecasts during the 1970s when the US economy had to face severe oil price shocks. But, coinciding with the Volcker disinflation period the Greenbook nowcasts performed worse for some time. Interestingly, during that period the one-year ahead forecasts document an informational advantage of the Greenbook forecasts over the SPF. Since about 1985, when the Great Moderation reduced overall volatility the forecast performance of both groups is rather similar. The tests show that since the mid-1980s there are no meaningful differences in the relative forecasting performance of inflation and output.

4.4 Testing for conditional predictive ability

An explanation for the superiority of central banks staffs' forecasts discussed in the literature is their better knowledge about the future interest rate path. Only few central banks share this information with the public in a systematic manner (see e.g. Sweden and Norway).⁶ But there are other shocks for which the degree of uncertainty faced by central bank staff and by private forecasters is similar (e.g. oil-price shocks, financial crises). Conditioning can help to improve the forecast selection given shocks common to both groups of forecasters. To test for superiority given such exogenous factors we use the conditional predictive ability (Wald type) tests proposed by Giacomini and White (2006). The test requires a sufficient number of observations and is therefore only applied to US data. It

⁶See also Rudebusch and Williams (2008) for a theoretical analysis.

provides information on whether changes in the relative forecasting performance of the Greenbook and the SPF are triggered by specific variables. We use a set of conditioning variables to proxy for uncertainty, data revisions, information about the interest rate, and oil⁷ and commodity prices.

First, we examine the relative forecasting performance for inflation and real GDP growth, given an uncertain economic environment. This is measured by the cross sectional dispersion for the quarterly forecasts⁸ (i.e. the dispersion of inflation, real GDP, industrial production, and housing starts). These variables are available from the Federal Reserve Bank of Philadelphia. Since the dispersion measures we use are obtained from the SPF they might include an SPF specific information disadvantage that central bankers are not subject to. To show that our results capture the impact of general economic uncertainty on forecast performance (rather than being spurious) we perform a robustness test. We use the predicted variance of inflation obtained from a simple GARCH(1,1) model instead of the dispersion of forecasts. Second, we check whether data revisions have an impact on the relative forecasting performance. We condition on the revisions in the variables inflation and real GDP growth. This set of revisions is constructed using the real-time data set from the Federal Reserve of Philadelphia. 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 performance. We use the absolute change of the Federal funds rate within the considered forecast horizon. Fourth, we condition on oil prices and the commodity price index. The null hypothesis is that given the information set Ω_t it is not possible to distinguished which forecast group has a lower error at horizon τ . It can be written as:

⁷Including oil prices is mostly due to recent findings — see European Central Bank (2012) — that imply a bias in ECB forecasts caused by oil price movements.

⁸The dispersion measure equals the 75th percentile minus the 25th percentile of the forecasts for quarter on quarter variables.

$$H_0 : E[L(\hat{f}_{t+\tau}^S) - L(\hat{f}_{t+\tau}^P)|\Omega_t] = 0. \quad (4)$$

Table 6 reports the results. First, if we condition the forecast performance on a given dispersion variable to proxy the uncertainty in the economic environment, we find that the central bank made better forecasts relative to the SPF in times of high uncertainty. This is the case for the nowcast of real GDP growth and for inflation for all forecast horizons, except for the nowcast. Hence, Greenbook forecasts for inflation over longer horizons are more accurate than private forecasts when overall uncertainty is high. But, the relative forecasting performance for output is not better when looking at longer horizons.

Second, if we condition the relative forecast performance on data revisions, which are calculated using the real-time vintages from the Federal Reserve of Philadelphia, we find that the relative forecasting performance is only significantly affected in the very short term. Surprisingly, revisions in inflation cause improvements in the relative forecasting performance for real GDP, and revisions in real GDP cause improvements in the relative forecasting performance for inflation. Though, as argued by Romer and Romer (2000), for most horizons the tests support the notion that the Fed staff makes better forecasts for reasons other than their early access to government statistics.

Table 6: Tests for conditional predictive ability

Horizon (quarters)	Uncertainty				Data revisions		Int. rate changes	Commodity prices	
	DPGDP	DRGDP	DIProd	DHousing	ARCH	RevPGDP			RevRGDP
US Inflation									
0	0.359	0.226	0.264	0.478	0.270	0.400	0.254	0.112	0.919
1	0.025	0.022	0.002	0.049	0.027	0.422	0.004	0.151	0.380
2	0.085	0.080	0.012	0.193	0.135	0.814	0.148	0.108	0.367
3	0.052	0.018	0.016	0.000	0.001	0.524	0.652	0.306	0.443
4	0.001	0.001	0.001	0.000	0.000	0.556	0.178	0.009	0.494
US GDP									
0	0.050	0.015	0.017	0.024	0.080	0.008	0.295	0.081	0.307
1	0.241	0.258	0.156	0.293	0.219	0.561	0.966	0.655	0.300
2	0.878	0.757	0.950	0.820	0.850	0.231	0.595	0.808	0.304
3	0.798	0.555	0.802	0.810	0.678	0.232	0.470	0.861	0.502
4	0.767	0.906	0.728	0.519	0.822	0.357	0.342	0.975	0.287

Notes: The table comprises p-values for the conditional predictive ability test by Giacomini and White (2006). P-values lower than 10 percent are marked bold. DPGDP, DRGDP, DIProd, and DHousing denote dispersion measures for the corresponding variables (GDP deflator, real GDP growth, growth of industrial production and housing starts). ARCH measures the volatility in past inflation rate and is used as an alternative uncertainty measure. RevPGDP and RevRGDP denote the latest revisions of the GDP deflator and real GDP growth. Int. rate changes of the Fed refers to quarter-on-quarter changes of 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.

Third, when testing for relative forecast performance conditional on future interest rate changes, the results seem to suggest that the Fed made better inflation forecasts when there were upcoming changes of the interest rate. As for the uncertainty factor this holds for longer horizon inflation forecasts and also for the nowcast of output, albeit the latter indicates only weak significance.⁹ That is, the Fed's inflation forecasts seem to have benefited from a better knowledge of future interest rate changes.

Fourth, when using oil prices and the HWWA index for energy, oil and raw materials, it turns out that these factors have no significant influence on the relative predictive ability of the competing forecasters.¹⁰ The test cannot reject the null, confirming the widespread notion that both groups 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

This paper sheds new light on the economics behind the finding by Romer and Romer (2000) that central bank staff forecasters outperform private forecasters. Since their seminal study was conducted around a decade ago, new data has become available and private forecasters have discovered the importance of making high quality forecasts of macroeconomic variables. Moreover, with the creation of a currency union in Europe, a new central banking system was founded, making it possible to test whether the predictions for the US also hold for the euro area.

⁹Since interest rate changes may be frequent in times of higher economic uncertainty, we check for correlation between these variables that may drive our results and, hence, distort our interpretation. However, we find only some correlation of the variables (0.30), implying that the results are mainly attributable to the separate effect of the interest rate.

¹⁰This finding is robust to using different measures for oil. Since the HWWA index also comprises commodity prices we only report the results for this measure in Table 6.

We present evidence based on euro area data over the last decade suggesting that staff outperforms private forecasters in the case of short-term forecasts of these variables. This is not the case for forecasts at longer horizons which are typically emphasized more in forward-looking monetary policy assessments. The US evidence presented in this paper confirms two seemingly competing findings in the literature: (i) that Greenbook forecasts outperform private forecasts on inflation and output and (ii) that this excess performance has deteriorated in past years. This assessment inspired us to look closer into the dynamics and the possible reasons for different forecast performance of both central bank staff and private forecasters. Using conventional forecast rationality tests in a moving window framework, we identify several periods for which the information efficiency of the forecasts disappears for both the Fed and the SPF. During the Great Moderation neither private nor staff forecasts are successful in explaining the variation of inflation around its mean, causing the rationality tests to reject in samples that mostly cover this period. Rationality of the Greenbook inflation forecasts is strongly rejected in many samples that include the Volcker disinflation period. This observation is not mirrored by the same extent in private forecasts and raises the issue why during this episode staff forecasts did not fully exploit all available information. Like in the US, the rationality tests reject rationality in the last part of the Great Moderation in the euro area. During the financial crisis volatility increased dramatically. Partly, this represented noise, causing large forecast errors and wide confidence bands. However, the volatility in the driving forces of inflation created new information that could be exploited in forecasting. The private forecasters appear to have had more problems with it than the ECB staff for which rationality is no longer rejected in that period. Moreover, in line with the findings of the previous literature on US data, additional tests show that both ECB staff and Fed

staff forecasts include information that would be valuable to private forecasters (at least for short term forecasts). For the ECB longer forecasts horizons were also included in the analysis. Here, we find that the relative advantage in forecasting is not present any more, possibly because private forecasters can include information contained in ECB staff forecasts by including the latest vintage of these forecasts in their model. Here, a comparison with US data from the SPF would have been interesting, but these are only available for 4 quarters ahead and the obvious alternative Blue Chip forecasts which has a longer horizon, is not fully comparable with the euro area SPF.

For the US, where sufficient long data for additional tests is available, a look into relative forecast performance - using relative forecast performance stability tests - indicates, that the importance of this information advantage changed substantially over time. We find that the differences in the forecasting performance can be explained by factors such as uncertainty and future interest rate changes. Especially the latter finding stresses the importance of the knowledge of the central bank's reaction function for which staff forecasts may include useful information. Several questions arise in the context of the present analysis and could be addressed in more depth in future research. When monitoring central bank forecasts do private forecasters fully understand the implications of the underlying technical assumptions, including those for the future interest rate path, for forecasting performance? Should a central bank refrain from publishing its staff forecast? Our paper seems to suggest that a narrowing of the gap between staff forecasts and private forecasts has happened on a more global scale. Since by publishing forecasts, this information becomes public knowledge, it could become apparent to the public that the central bank is more uncertain about the future outlook than it would otherwise admit. This in turn would make it more difficult for the central bank to convince

the public it makes the right policy.

The findings in this paper suggest that nowadays communications by the Fed and the ECB aiming at influencing long-term inflation expectations can hardly rely on an information advantage in their forward-looking assessment of future inflation or growth perspectives. That is, other factors including clarity about the price stability goal, a monetary policy strategy, credible and timely policy actions, monetary policy transparency, and clear communications on monetary policy become more important elements in stabilizing inflation expectations. Furthermore, even if a central bank has an information advantage in forecasting inflation and output, it is not clear whether it should exploit forecasts in its communication to guide markets. In view of Goodhart's law it can be expected that attempts to do so might become self-defeating as soon as they destabilise the relation between inflation forecasts and inflation, thereby reducing the credibility of the central bank. However, since we are able to identify specific factors explaining the relative forecast performance, it seems unlikely that this is already the case.

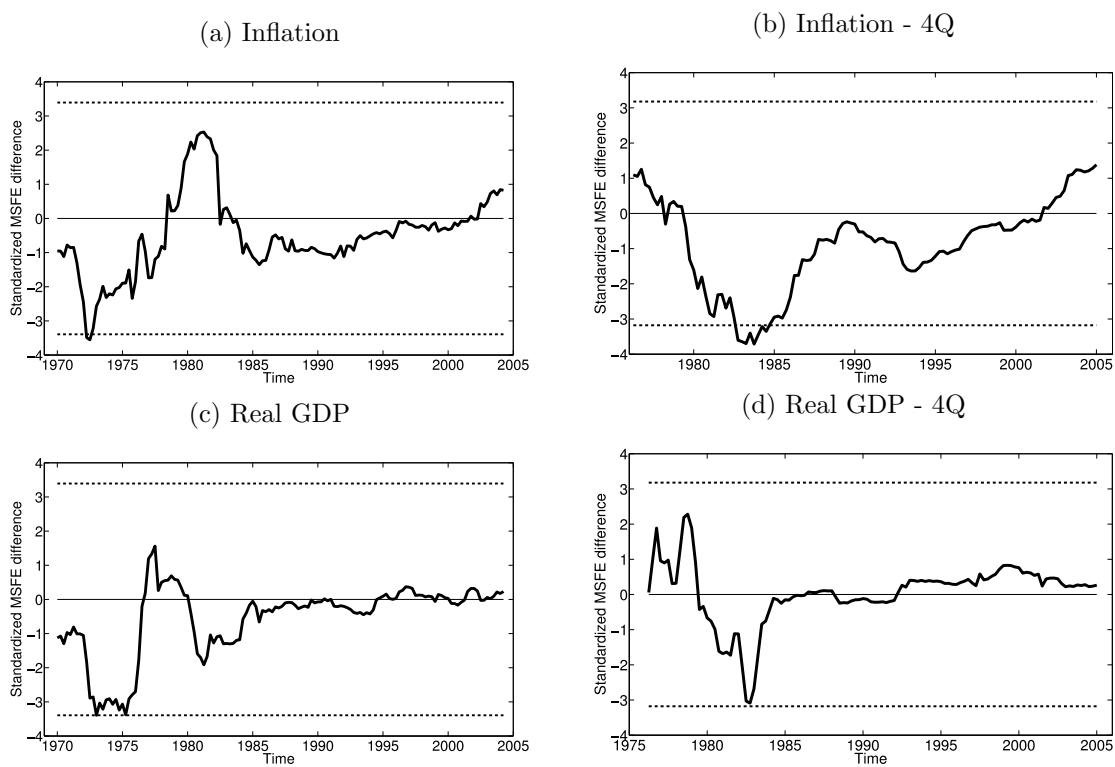
References

- Andrews, D. W., Lee, I. and Ploberger, W. (1996). Optimal changepoint tests for normal linear regression, *Journal of Econometrics* **70**(1): 9–38.
- D’Agostino, A. and Whelan, K. (2008). Federal Reserve information during the Great Moderation, *Journal of the European Economic Association* **6**(2-3): 609–620.
- Ellison, M. and Sargent, T. J. (2009). A Defence of the FOMC, *Technical Report DP7510*. CEPR Discussion Paper.
- European Central Bank (2012). Monthly Bulletin, (06/2012).
- Gamber, E. N. and 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. and Mandal, R. J. (2003). Evaluating FOMC forecasts, *International Journal of Forecasting* **19**(4): 655–667.
- Genre, V., Kenny, G., Meyler, A. and Timmermann, A. (2010). Combining the Forecasts in the ECB Survey of Professional Forecasters: Can Anything Beat the Simple Average?, *Working Paper Series 1277*, European Central Bank.
- Giacomini, R. and Rossi, B. (2010). Forecast comparisons in unstable environments, *Journal of Applied Econometrics* **25**(4): 595–620.
- Giacomini, R. and White, H. (2006). Tests of conditional predictive ability, *Econometrica* **74**(6): 1545–1578.
- Giannone, D., Henry, J., Lalik, M. and Modugno, M. (2010). An Area-Wide Real-Time Database for the Euro Area, *Working Paper Series 1145*, European Central Bank.
- Goodfriend, M. (1997). Monetary policy comes of age: a 20th century odyssey, *Federal Reserve Bank of Richmond Economic Quarterly* **83**(1): 1–22.
- Hubert, P. (2012). ECB Projections as a Tool for Understanding Policy Decisions, *Technical report*.
- Kenny, G. and Morgan, J. (2011). Some lessons from the financial crisis for the economic analysis, *Occasional Paper Series 130*, European Central Bank.
- McCracken, M. (2010). Disagreement at the FOMC: the dissenting votes are just part of the story, *The Regional Economist* pp. 10–16.
- Moutot, P., Jung, A. and Mongelli, F. P. (2008). The Workings of the Eurosystem, Monetary Policy Preparations and Decision-Making - Selected Issues, *Occasional Paper Series 79*, European Central Bank.

- Peek, J., Rosengren, E. S. and Tootell, G. M. B. (2003). Does the federal reserve possess an exploitable informational advantage?, *Journal of Monetary Economics* **50**(4): 817–839.
- Reifschneider, D. and 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. and Romer, D. H. (2000). Federal Reserve information and the behavior of interest rates, *American Economic Review* **90**(3): 429–457.
- Romer, C. D. and Romer, D. H. (2008). The FOMC versus the Staff: Where can monetary policymakers add value?, *American Economic Review: Papers & 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. and Sekhpoysan, T. (2011). Forecast Optimality Tests in the Presence of Instabilities, *Technical report*. Working Paper.
- Rudebusch, G. D. and Williams, J. C. (2008). Revealing the secrets of the temple: the value of publishing central bank interest rate projections, *Asset Prices and Monetary Policy*, NBER Chapters, National Bureau of Economic Research, pp. 247–289.
- Sims, C. A. (2002). The role of models and probabilities in the monetary policy process, *Brookings Papers on Economic Activity* **2002**(2): 1–40.
- Stock, J. H. and Watson, M. W. (1993). A simple estimator of cointegrating vectors in higher order integrated systems, *Econometrica* **61**(4): 783–820.
- Stock, J. H. and Watson, M. W. (2007). Why has US Inflation become harder to forecast, *Journal of Money, Credit and Banking* **39**(s1): 3–33.
- 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.

A Figures

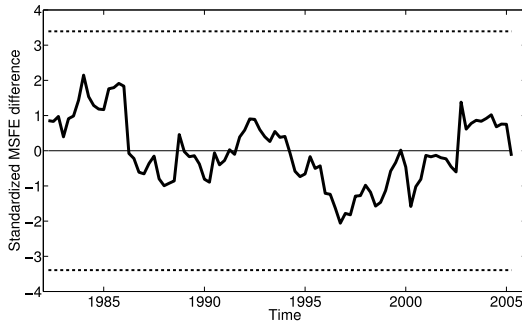
Figure 1: Fluctuation test statistic - United States



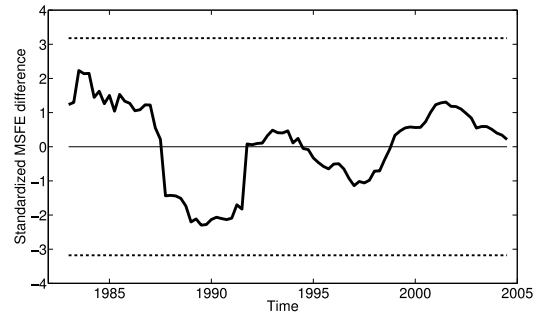
Notes: The solid line shows the fluctuation test statistic and the dashed lines represent the corresponding critical values.

Figure 1 (cont.): Fluctuation test statistic - United States

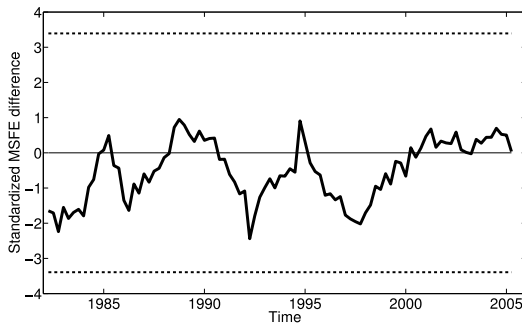
(e) Real Consumption



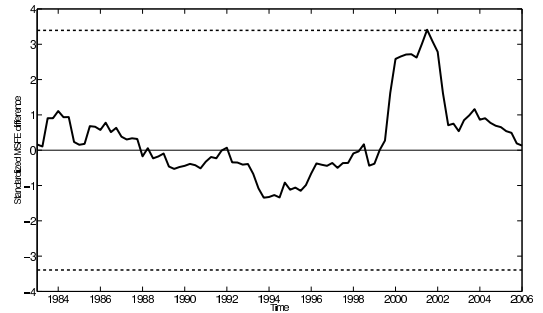
(f) Real Consumption - 4Q



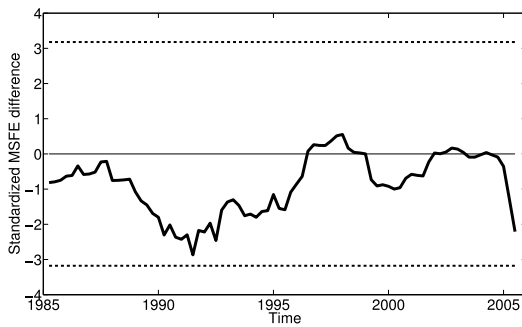
(g) Real Fixed Business Investment



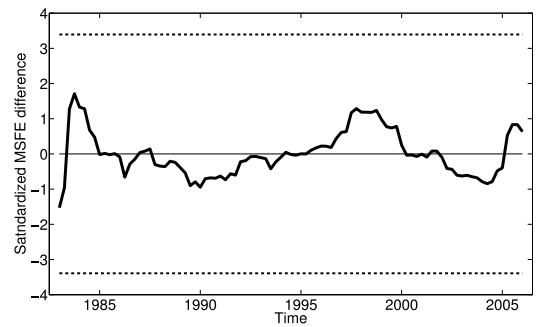
(h) Real Fixed Business Investment - 4Q



(i) Residential Investment



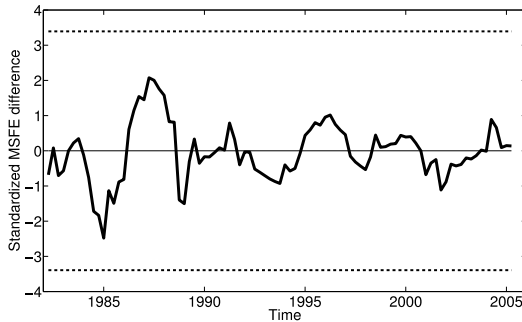
(j) Residential Investment - 4Q



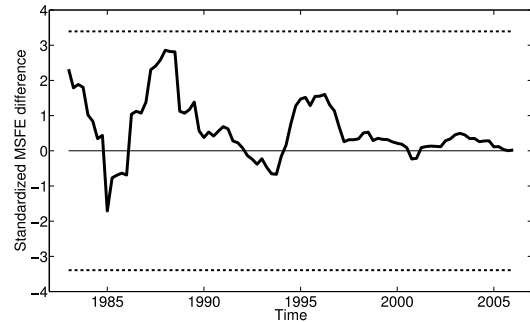
Notes: The solid line shows the fluctuation test statistic and the dashed lines represent the corresponding critical values.

Figure 1 (cont.): Fluctuation test statistic - United States

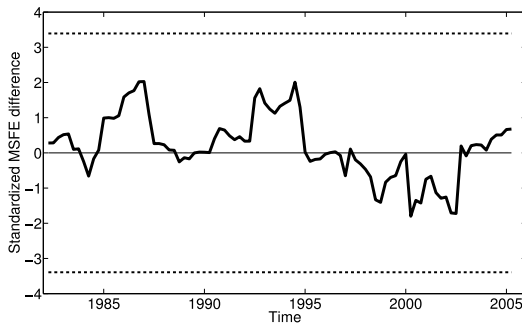
(k) Federal Government Cons.



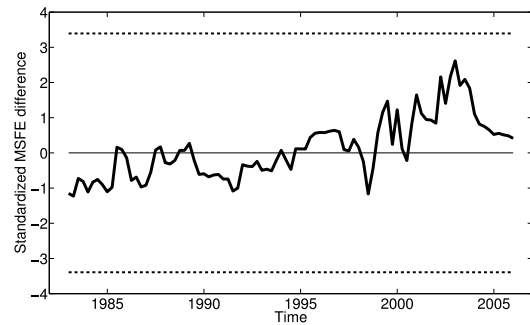
(l) Federal Government Cons. - 4Q



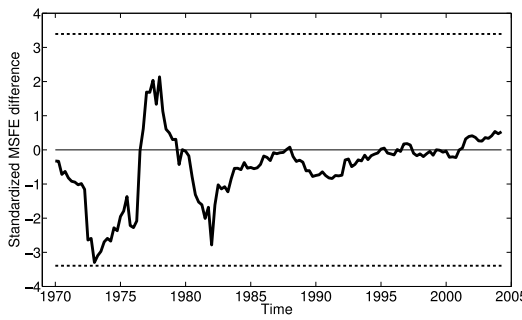
(m) Local and State Gov. Cons.



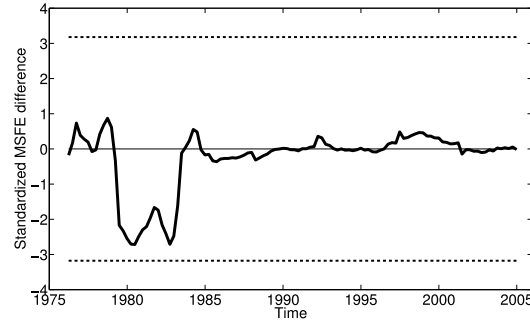
(n) Local and State Gov. Cons. - 4Q



(o) Nominal GDP

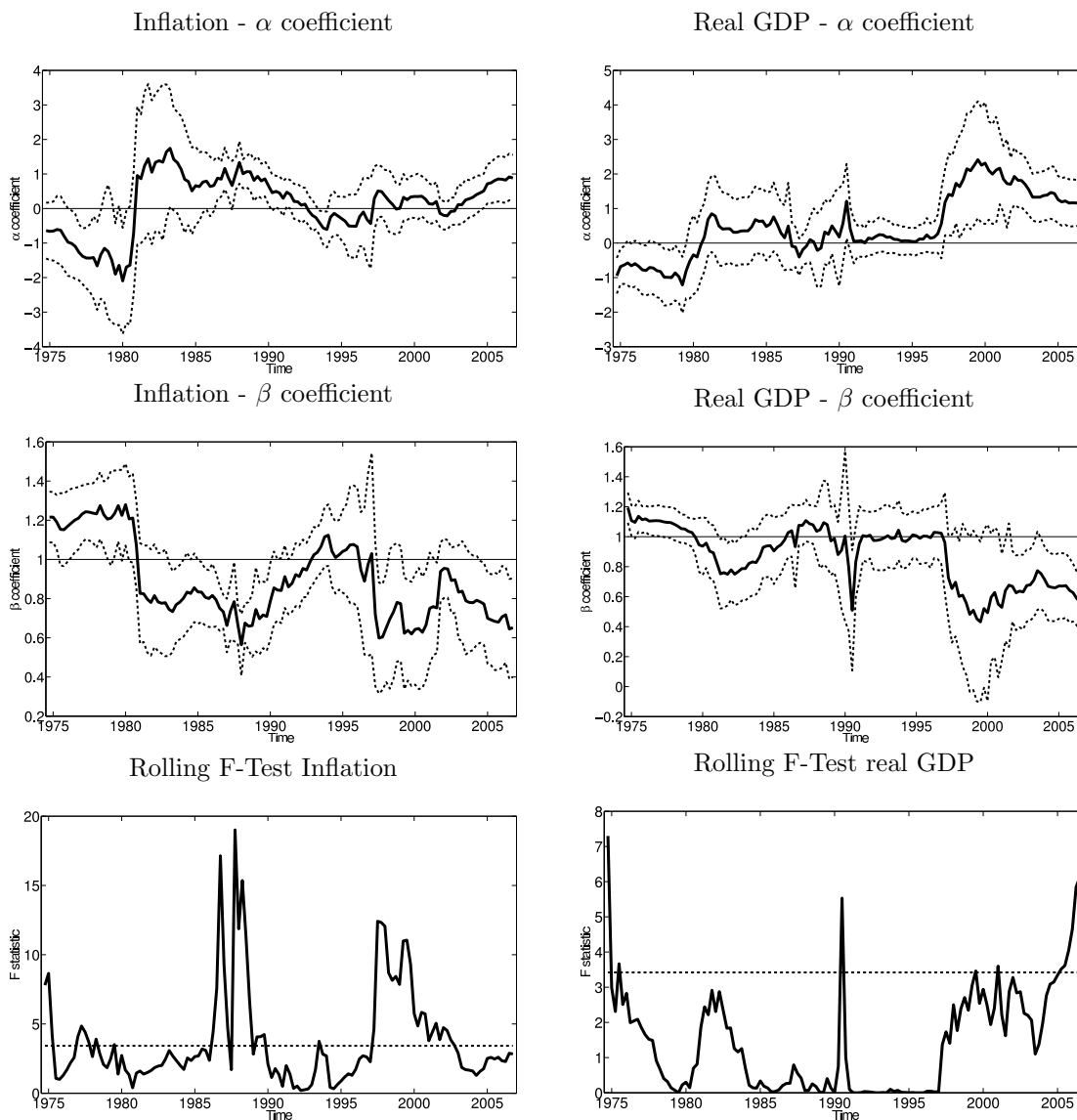


(p) Nominal GDP - 4Q



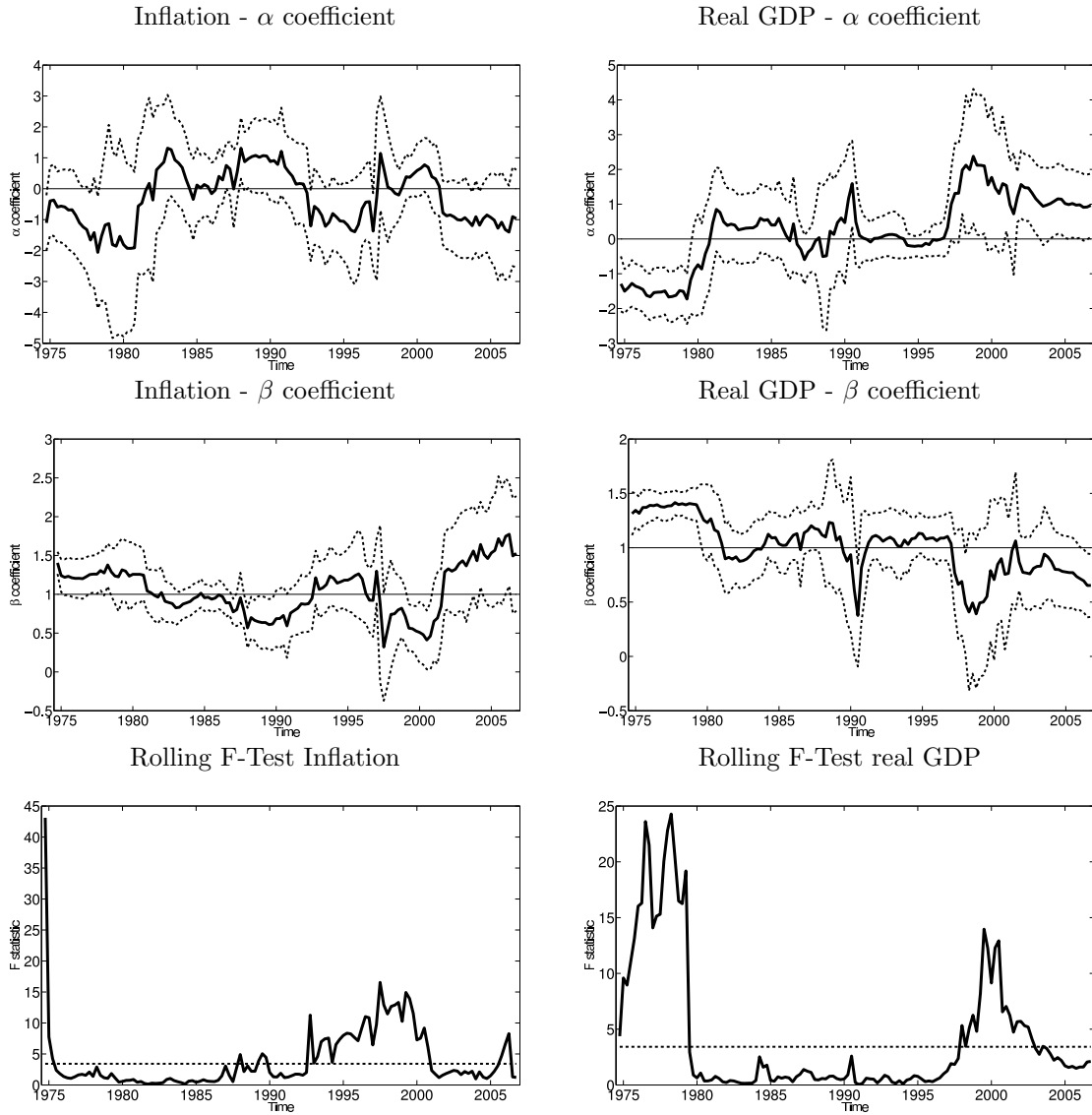
Notes: The solid line shows the fluctuation test statistic and the dashed lines represent the corresponding critical values.

Figure 2: Rolling Window Estimation - Federal Reserve



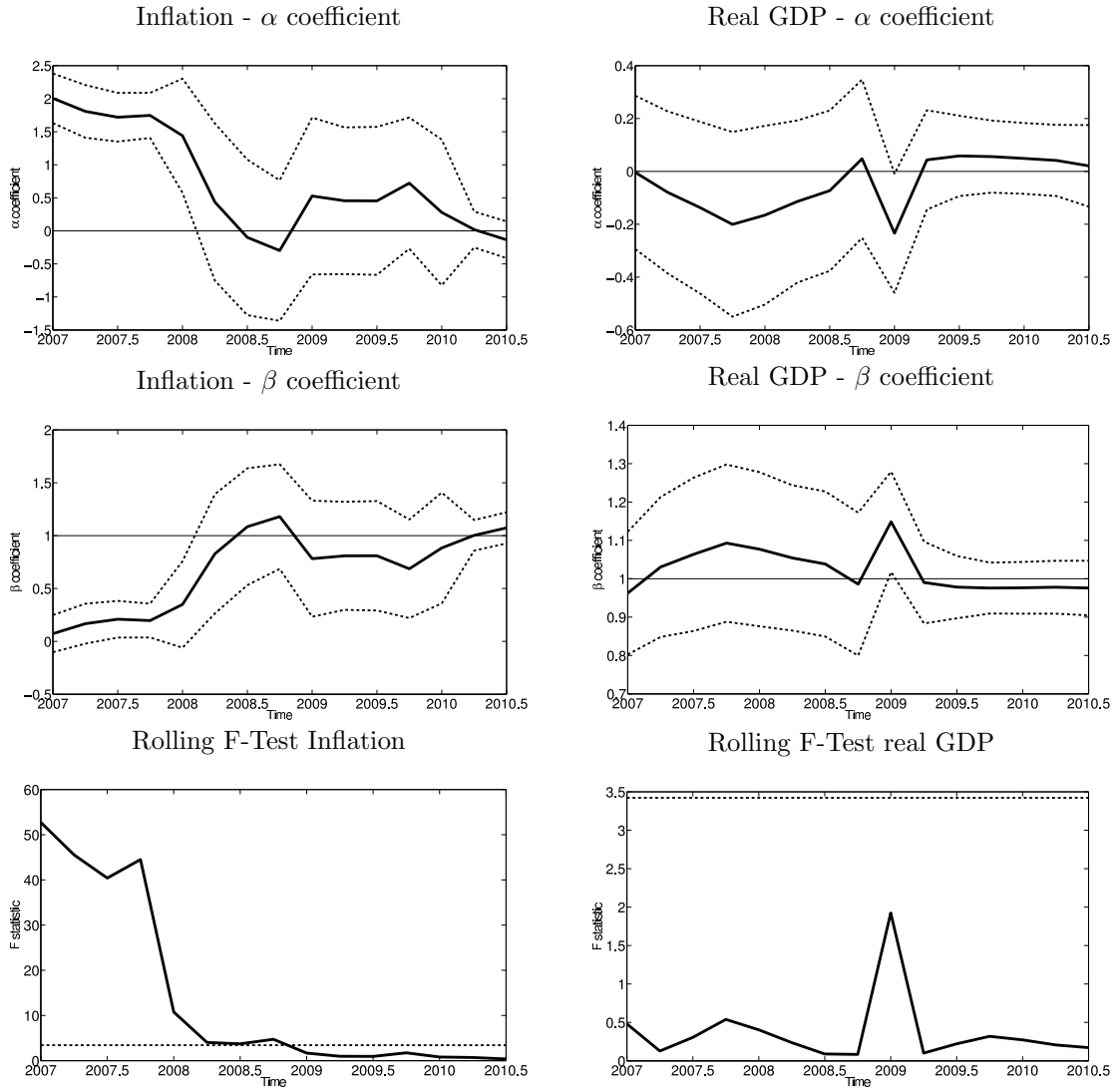
Notes: The top four pictures show rolling window estimates for the individual α and β coefficients (solid lines) with corresponding 95% confidence bounds (dashed lines). The window size for estimation comprises 25 observations. The two bottom pictures show the corresponding evolution of the F-statistic (for the joint hypothesis $\alpha = 0$ and $\beta = 1$).

Figure 3: Rolling Window Estimation - US SPF



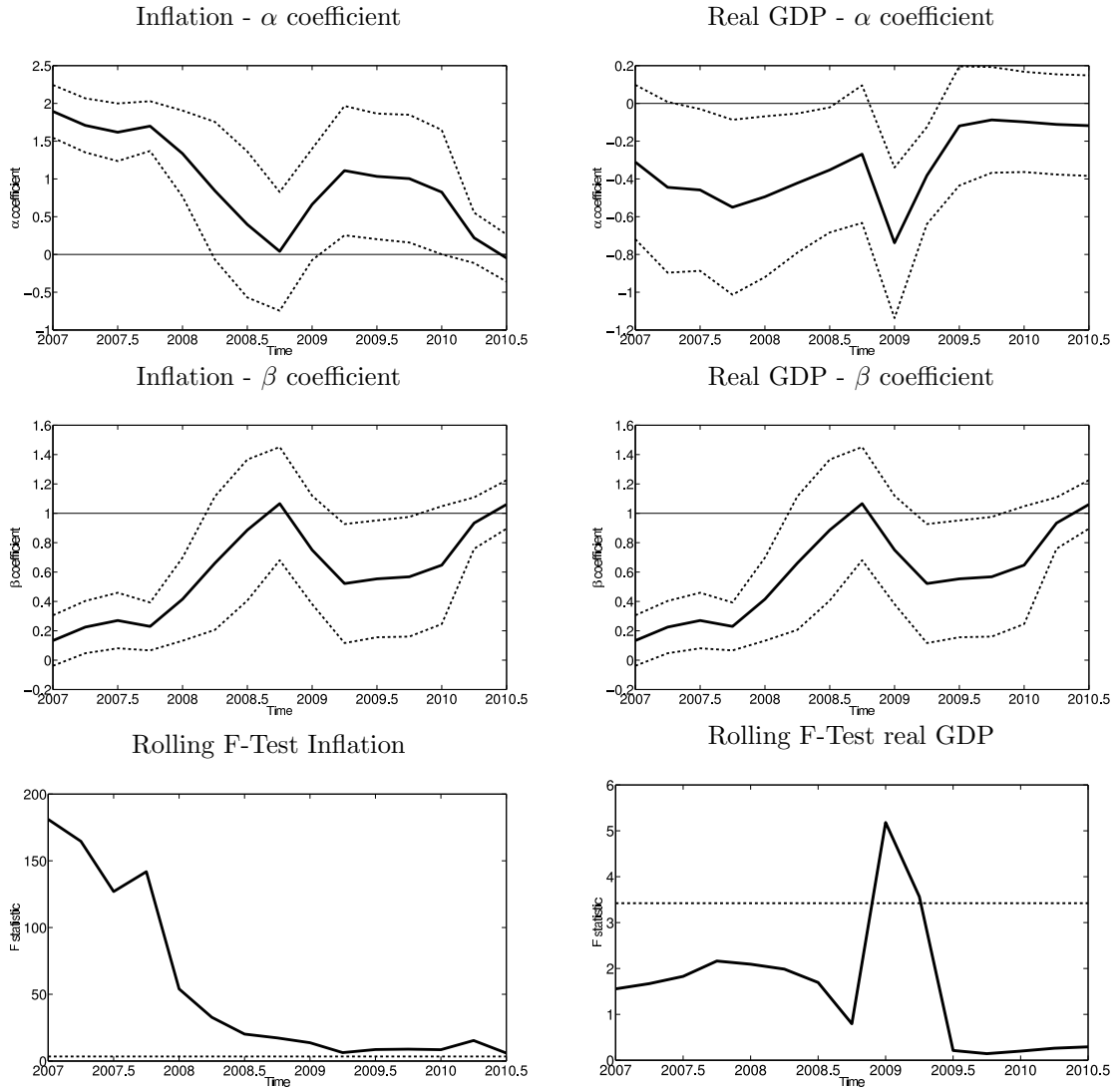
Notes: The top four pictures show rolling window estimates for the individual α and β coefficients (solid lines) with corresponding 95% confidence bounds (dashed lines). The window size for estimation comprises 25 observations. The two bottom pictures show the corresponding evolution of the F-statistic (for the joint hypothesis $\alpha = 0$ and $\beta = 1$).

Figure 4: Rolling Window Estimation - ECB



Notes: The top four pictures show rolling window estimates for the individual α and β coefficients (solid lines) with corresponding 95% confidence bounds (dashed lines). The window size for estimation comprises 25 observations. The two bottom pictures show the corresponding evolution of the F-statistic (for the joint hypothesis $\alpha = 0$ and $\beta = 1$).

Figure 5: Rolling Window Estimation - Euro Area SPF



Notes: The top four pictures show rolling window estimates for the individual α and β coefficients (solid lines) with corresponding 95% confidence bounds (dashed lines). The window size for estimation comprises 25 observations. The two bottom pictures show the corresponding evolution of the F-statistic (for the joint hypothesis $\alpha = 0$ and $\beta = 1$).