

Jung, Alexander; El-Shagi, Makram; Giesen, Sebastian

Conference Paper

Does Central Bank Staff Beat Private Forecasters?

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: Business Cycles I, No. A02-V2

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Jung, Alexander; El-Shagi, Makram; Giesen, Sebastian (2013) : Does Central Bank Staff Beat Private Forecasters?, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: Business Cycles I, No. A02-V2, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/79925>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Does Central Bank Staff Beat Private Forecasters?*

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

February 19, 2013

Abstract

This paper assesses the relative performance of central bank staff forecasts and of private forecasters for inflation and output. We show that the Federal Reserve (Fed), and less so the European Central Bank (ECB), has a significant information advantage concerning inflation and output forecasts. Using recently developed tests for conditional predictive ability and forecast stability for the US, we find that the driving forces behind the narrowing of the information advantage of Greenbook forecasts have coincided with the Great Moderation.

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, an anonymous referee, 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.europa.eu.

1 Introduction

Today, 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. The challenge of making good forecasts throughout the financial crisis has contributed to renewed interest in the question whether central bank staff makes better forecasts of the macroeconomy than private forecasters. Since the outbreak of the financial crisis in 2007 volatility has increased. Owing to the extraordinary volatility and uncertainty, the forecast performance of most central banks, and of private forecasters has significantly deteriorated (see Kenny and Morgan (2011)).

Anecdotal evidence supports this point. First, in the presence of persistent oil price shocks, several inflation-targeting central banks underpredicted inflation for some time. Second, Federal Reserve (Fed) chairman Bernanke mentioned in a press conference (on 12 December 2012) that the Fed overpredicted real GDP growth in past years. Third, another case in point is the Stockton (2012) Report which suggests that the Bank of England's recent forecast performance has deteriorated and has been "marginally worse" than that of private forecasters. This finding appears to be in contrast with Romer and Romer (2000) who demonstrated for the US case that internal macroeconomic forecasts by Fed staff outperform private forecasters based on their better forecasting expertise.

Some researchers have observed changes in the relative forecasting ability of inflation and output. For example (D'Agostino and Whelan (2008), Gamber and Smith (2009) and Rossi and Sekhpoysan (2011)) find that the Fed's

information advantage over private actors disappeared in the course of the Great Moderation. This led to a controversy on whether central bank staff performs better than private forecasters, when uncertainties are elevated. Tulip (2009) argues that inflation and output have become more predictable over recent years and forecasters should perform largely similar. Other researchers (see Stock and Watson (2007)) argue that over time it has become more difficult to forecast inflation with the necessary degree of precision even though central banks better anchor inflation expectations and the inflation process has become less volatile.

Our paper contributes to the literature in the following ways. We compare staff forecasts for inflation and output with corresponding private forecasts for the US and the euro area. We find that the Fed's Greenbook forecasts of inflation and output used to provide a significant information advantage over private sector forecasts, but coinciding with the Great Moderation forecast performance has become more similar. Applying recently developed tests for conditional predictive ability and forecast stability (see e.g. Giacomini and White (2006) and Giacomini and Rossi (2010)) for the US, we find that significant changes in the relative forecast performance have occurred since the beginning of the 1990s. We are also able to econometrically identify the economic reasons for the development of relative forecast performance. For the euro area, empirical tests can only be applied to the period since the inception of the European Central Bank (ECB) in 1999. Our results suggest that private forecasters did not outperform Eurosystem/ECB staff forecasts for inflation and output. Some results tend to suggest that the ECB has

made better inflation forecasts than the private sector.

The remainder of the paper is structured as follows. Section 2 briefly reviews the literature. Section 3 explains the data used in this paper and provides some results on relative forecast accuracy. Section 4 separately analyses forecasting performance of the Fed's and Eurosystem/ECB's staff forecasts for inflation and output relative to those by private forecasters. It applies recently developed tests for conditional predictive ability and forecast stability to US data. Section 5 concludes.

2 A brief review of the literature

Since the mid-1990s, most inflation targeting central banks have regularly provided the public with their staff forecasts when explaining their monetary policy decisions. Also other central banks like the ECB regularly (usually quarterly) publish macroeconomic forecasts prepared by their staff. The Fed is somewhat different in that it releases the FOMC forecast made by policy-makers once a quarter, but its Greenbook staff forecasts for each FOMC meeting are only published with a lag of about five years. These staff forecasts are therefore not available to the public when assessing the Fed's monetary policy stance and the economic outlook. But, 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).¹

¹In a study, Romer and Romer (2008) look into the controversial question whether the published FOMC policy-maker forecasts are inferior or similar to internal Greenbook

Why should staff forecasts be superior to private forecasts? In fact, this is a puzzling proposition, because the level of data and model uncertainty is profound for all forecasters. A priori it is therefore not clear whether forecasts by central bank staff (or by international organizations such as IMF and OECD) should be more or less accurate than forecasts produced by the private sector (see Batchelor (2000)). One of the main reasons here is the high level of transparency achieved by main central banks over the last decade. Thereby, they have deliberately reduced their relative information advantage with the aim to enhance the effectiveness of monetary policy (see Woodford (2005)). In parallel to central banks' efforts to increase transparency about their economic assessment, the private sector (mainly banks and other agencies) has increased the amount of resources it invests in making their own forecasts of these variables. It has allowed them to scrutinize the central banks' forward-looking assessment, when forming expectations, and it has contributed to better predictions of future monetary policy decisions (see Brand, Buncic and Turunen (2010) and Blattner, Catenaro, Ehrmann,

forecasts prepared by staff, which are not available in real-time. They find that FOMC forecasts do not provide useful information relative to the Greenbook forecast even though FOMC members know the staff forecast when making their individual forecast. A study by Ellison and Sargent (2009) suggests that an inferiority of FOMC forecasts is at odds with evidence documenting that differences between FOMC and Greenbook forecasts are very small. We cannot resolve this debate here, but emphasise that the following important differences between both types of forecasts could have a bearing on the relative forecast performance. First, the FOMC forecast is made by individual FOMC members, it is not a staff forecast and it has not yet been established as a consensus forecast that is based on identical interest rate assumptions across FOMC members (currently work is under way in that direction). The accuracy of FOMC forecasts could be negatively influenced by specific factors including strategic motives by individual members and the non-harmonised interest rate assumption (see McCracken (2010) and Tillmann (2011)). Second, it has a larger dispersion around the mean, because the range of possible outcomes considered by all individual policy-makers given multiple uncertainties may be more disperse than what Fed staff forecasts report.

Strauch and Turunen (2008)). Lange, Sack and Whitesell (2003) find increased predictability of FOMC decisions because of improved transparency. In a similar spirit, Swanson (2006) suggests that since the late 1980s increases in Federal Reserve transparency have been instrumental to the ability of both US financial markets and the private sector to forecast the federal funds rate at horizons of several months. Still, in the literature it is widely assumed that central banks have an information advantage relative to the private sector, which could contribute to a better forecast accuracy in relative terms. First, they have more timely and complete knowledge of official statistics and may have access to first estimates of data releases. Second, they have more insight into their likely reaction to future shocks and, in the absence of forward guidance, they should have better knowledge on their own intentions on future interest-rate setting.²

As institutions funded by the government, central banks could possibly be subject to political pressure in their response to economic shocks and this may diminish the accuracy of their forecasts. At the very least, political pressure could imply a deterioration of the quality of forecasts and would become visible in frequent revisions of first estimates. Today, most central banks are independent in the pursuit of their monetary policy goals, though to a varying extent (see Alesina and Summers (1993) and the more recent evidence provided in Moutot, Jung and Mongelli (2008)). Central bank independence

²In the case of international organizations, this information advantage may be attenuated by the fact that forecasters are not based in the countries they are forecasting. They may benefit from consultations with central bank staff during country missions, but may not have access to various informal pieces of information which is available to local forecasters regardless of whether they work for the central bank or the private sector (see Abreu (2011)).

ensures in most instances that central banks are in a position to make an independent assessment of the monetary policy stance and the underlying assumptions.

As concerns inflation forecasts by central bank staff (and possibly for output forecasts, if they are made consistent with the inflation forecast), a further argument is that in the pursuit of a price stability goal central banks could have the incentive to align forecasts at the policy horizon with their numerical inflation target. While this could help them to stabilize private expectations and thereby enhance the effectiveness of monetary policy, it would result in a deterioration of forecast accuracy in absolute terms and relative to other forecasters. For example, assessments of inflation forecasts by the Bank of England and the Swedish Riksbank provide some evidence on the presence of this kind of bias, when real-time forecasts are compared with their outcomes (see e.g. Jung (2011)).

Most central banks are only responsible for preparing statistics on money and interest rates, and other government institutions are in charge of preparing the National Accounts and the fiscal data. Even, if central bank independence is considered to be a strong argument in this discussion, it still could be that the other government institutions, which provide important inputs to forecasting exercises are subject to political pressure (see Frankel and Schreger (2012)). These institutions may have incentives to provide data which cast a favorable light on the government and thereby contribute to worsen forecast accuracy of those who use these data. In the literature, anecdotal evidence demonstrates the relevance of this kind of behavior. For example, in the US the Congressional Budget Office (CBO) (1996) has shown

that its own independent forecasts for the US economy were comparable to the Blue Chip forecasts, but consistently more accurate than those of the politically influenced Office of Management and Budget. At the same time, it must be acknowledged that statistical agencies all over the world make continuous efforts to improve the timeliness and quality of their statistics. Moreover, private forecasters often have no other sources when making their macroeconomic forecasts, but given knowledge about quality problems of official statistics, they may more easily justify the use of other non-official sources as inputs. This could have a positive impact on their forecast accuracy.

The performance of private sector forecasts relative to staff forecasts has been the subject of a series of empirical studies. For reasons related to data availability, most studies examining this issue have been made for the US. 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.

Sims (2002) finds that 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 move-

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			
Hubert (2012)	MFE, RMSFE	1999 - 2009	Pooling private forecasts can improve inflation forecasts, but not output forecasts
Genre, Kenny, Meyler and Timmermann (2013)	RMSFE	2004 - 2011	Eurosystem 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.

ments 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.

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 Romer and Romer (2000), Sims (2002), Gavin and Mandal (2003), 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 to these studies, Reifschneider and Tulip (2007) find that Greenbook forecasts since 1986 are not more accurate than private forecasts. Studies by Gamber and Smith (2009) and Rossi and Sekhpoysan (2011) for an extended sample up to 2006 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. Since the mid-1980s, private forecasters have continuously improved their forecasting methods and many of them are applying state of the art forecasting models. Private research in-

stitutes also spend vast resources to make high quality forecasts and often publish these forecasts. In addition, the presence of information cascades (see Bikhchandani, Hirshleifer and Welch (2008)) has contributed to a better sharing of information and to a reduction in the dispersion of private forecasts.

For the euro area the empirical literature on forecast comparison is scant. In his analysis of the effect of the Eurosystem/ECB's forecasts on private actors, Hubert (2012) includes a brief assessment on relative forecast accuracy. Based on simple measures (RMSFEs) he concludes that the Eurosystem/ECB's short-term inflation forecasts outperform private forecasts. Genre, Kenny, Meyler and Timmermann (2013) make a different argument. They show that even in the difficult circumstances of a financial crisis there is the potential to improve forecast performance by pooling private forecasts from the ECB Survey of Professional Forecasters (SPF).

When comparing the evidence for the Fed and for the ECB, it should be borne in mind that the empirical tests usually require longer time series, which are only available for the Fed since the ECB was founded in 1999. In view of data availability constraints, it is not possible to directly compare the results for the US and the euro area in a meaningful way. Such a comparison would require a similar coverage of the data for both currency areas, which is currently only possible for a short period of about six years. Rossi and Sekhpoysan (2011) show that there could be considerable instabilities over time so that the results may be sample dependent. Croushore (2012) cautions against reading too much into results obtained for certain sub-samples. Also for conceptual reasons, it may turn out to be difficult to directly compare

evidence on relative forecast performance for the US and the euro area.

On the one hand, there are conceptual similarities between both central banks' forecasting procedures, which could suggest that the forecast performance relative to the private forecasters should be broadly similar. First, the forecasting process at both central banks involves a wide range of staff and several econometric models. Though, at the Eurosystem the number of experts involved is higher, because also experts from the National Central Banks (NCBs) are involved in it (see Moutot et al. (2008)). Second, the forecasting process is informed by a model and includes a substantial part of expert judgement. Third, both central banks internally track their staff forecast performance relative to other forecasters with the aim to improve forecasting accuracy and methods.

On the other hand, there are conceptual differences between the Eurosystem/ECB's staff forecast and the Fed's Greenbook forecast which per se may explain a difference in relative forecast performance with private forecasters. First, the Eurosystem regularly publishes its staff forecast, whereas the Greenbook forecast is not available to private forecasters in real-time. Second, different interest rate assumptions have been used. Fed staff conditions its forecasts on appropriate monetary policy, but it has remained vague about this assumption. The ECB initially conditioned its forecasts on constant interest rates and since June 2006 it has used a market interest rate assumption.

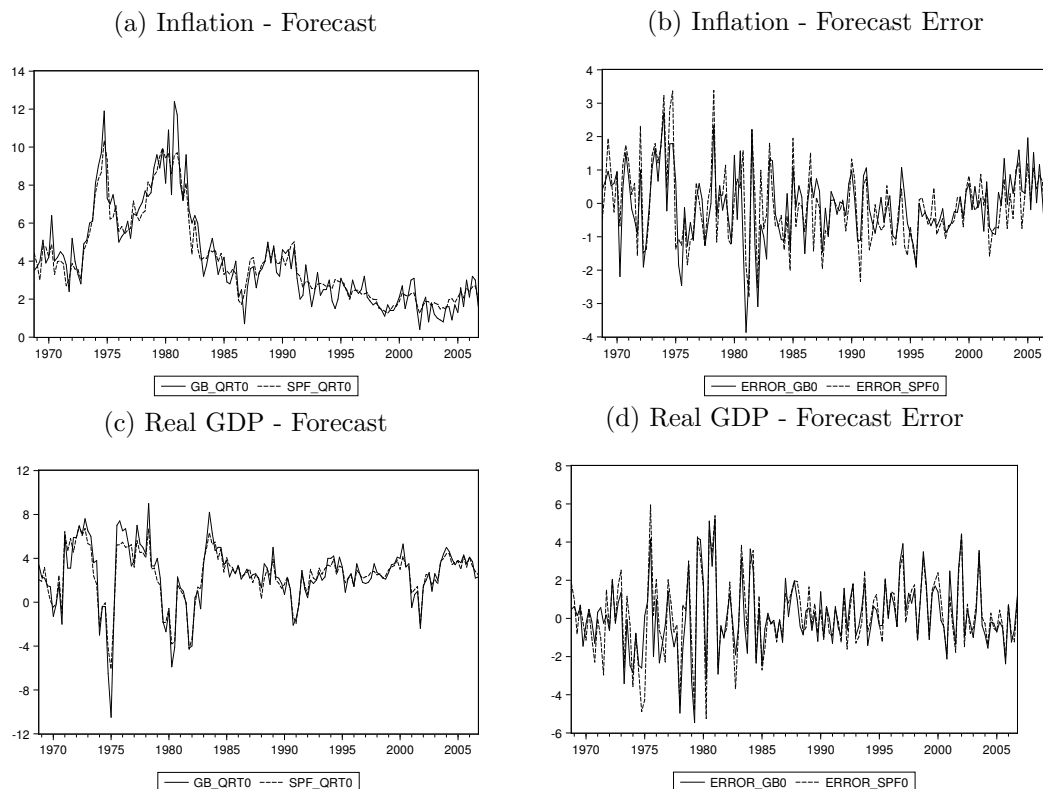
3 Data and relative forecast accuracy

In order to assess the empirical properties of forecasts provided by different institutions, it is necessary to set up a data base with real-time information on the various forecasts and their outcomes.³ This paper uses quarterly forecasts for inflation and (real) output from the Greenbook forecast, the Eurosystem/ECB staff forecasts and Surveys of Professional Forecasters (SPF) for the US and the euro area. We include available Greenbook forecasts for the sample 1968Q4 to 2006Q4 as well as Eurosystem/ECB staff forecasts for the sample 2000Q4 to 2012Q1. To proxy private forecasters in both economies, we include data from the respective Survey of Professional Forecasters for the corresponding sample. In the following we document data used separately for the United States and for the euro area.

For 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 of the Federal Reserve Bank of Philadelphia. These data 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

³There have been several severe revisions of the data during our sample period. In particular, the revisions of GDP in the US have been so strong that forecasts from the early part of our sample contain almost no information about today's view on GDP growth at the respective date. Therefore, for US GDP most of our results do not hold when using the final vintage of data instead of real time data. The results for US inflation, where the revisions have been less substantial, are qualitatively robust to replacing real time data by final data. For the euro area our analysis covers a more recent period. The last vintage of data is fairly similar to real time data; and results are similar with both first and final vintage data. All results reported in the remainder of the paper use real time – i.e. first vintage – data.

Figure 1: Forecasts and forecast errors (for the current quarter) for the US



Governors for each FOMC meeting. Greenbook forecasts are available for five or six quarters into the future, though the horizon of the forecast varies over time and with the date of the FOMC meeting. They generally report forecasts in terms of real GDP growth, but before 1992 the reported data 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 are used more prominently in policy debates.

The SPF is the oldest quarterly survey of macroeconomic forecasts in the

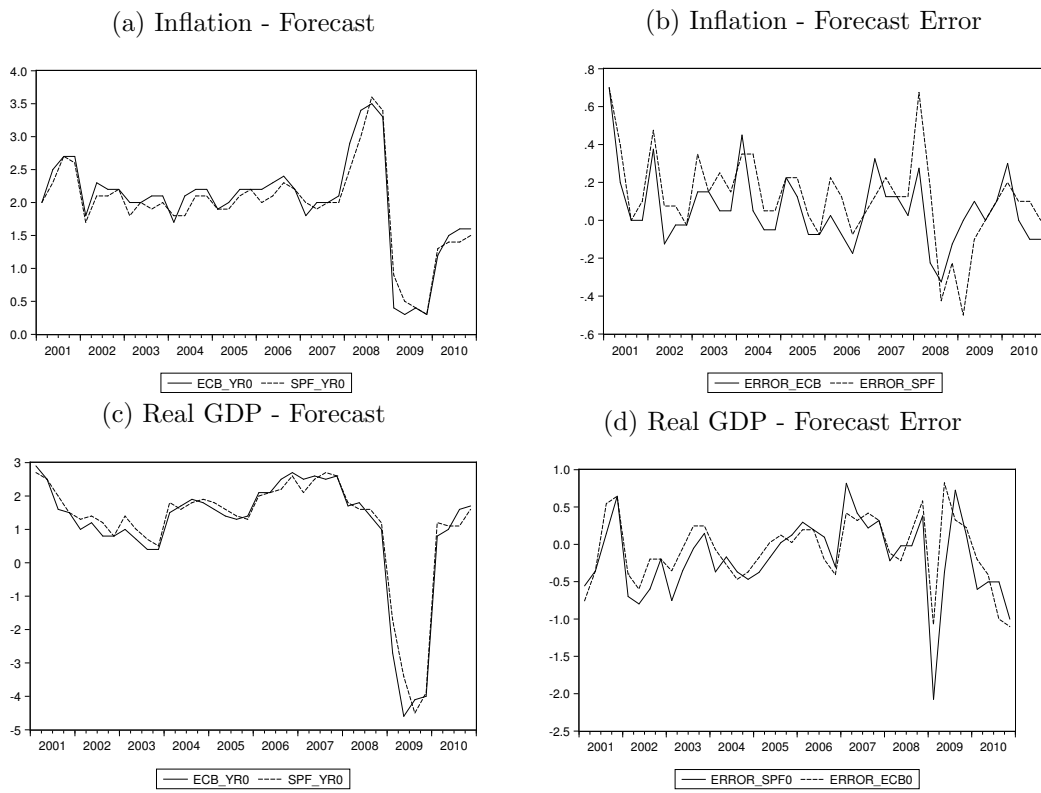
US. It has been conducted by the American Statistical Association and the National Bureau of Economic Research. The Bank of Philadelphia took over the survey as of 1990 and as of 1990Q3 has transformed it into a real-time survey. Forecasts for core CPI inflation, PCE inflation, and core PCE inflation were only added as late as 2007Q1. We therefore use the GDP (GNP) deflator (and, where meaningful we also report tests for CPI inflation). For real GDP (GNP) observations are fully comparable with the Greenbook forecasts.

Concerning the timing the Greenbook forecasts are released to the FOMC members prior to each meeting (the time of the month when the forecast is made also varies, because the date of the FOMC meeting varies). For the purpose of this study, we use the observations that become available in the following months: March, June, September and December. The timing of the SPF is such that new observations are released to the public by middle or end of the following months: February, May, August and November.⁴ In our regressions, we use forecasts which are made in the same quarter. When comparing the forecasts from the Greenbook with the SPF forecasts we hence put the SPF forecasters at a slight timing disadvantage over the Federal Reserve Board staff. Figure 1 compares the Greenbook nowcasts with the SPF nowcasts for inflation and output. It also reports corresponding forecast errors. The Greenbook and SPF forecasts and the corresponding forecast errors are highly correlated for the whole sample starting in 1970. It indicates generally similar forecasting strategies of both types of forecasters. Table 2

⁴For details see documentation of the Fed Philadelphia: <http://www.phil.frb.org/research-and-data/real-time-center/survey-of-professional-forecasters/spf-documentation.pdf>.

reports the mean square forecast errors (MSFE) for the forecasts plotted in the figures for several horizons. This comparison of forecast accuracy shows that the SPF forecasts for inflation and output are inferior to the Greenbook forecasts at all horizons considered.

Figure 2: Forecasts and forecast errors (for the current quarter) for the Euro area



For the euro area, we use the published Eurosystem staff macroeconomic projections for inflation and output in real time as well as the corresponding projections from the SPF (source: ECB website). The projections in June and December are prepared by Eurosystem staff and the projections in

March and September are prepared by ECB staff. The ECB publishes staff projections for inflation (HICP) and real GDP growth for the euro area as a whole, both for the current year and the next year (and in December for two years ahead). The precise horizon of the forecast varies with the date of its release. The ECB reports ranges rather than point forecasts. To allow for a comparison with private forecasters we proxy point forecasts by the midpoint of the provided range. This does not necessarily reflect the precise underlying point forecast of the staff. Additional breakdowns of GDP components are provided by the ECB, but for these variables no SPF measure is available for comparison.

The ECB Survey of Professional Forecasters (ECB SPF) provides quarterly data on private expectations about inflation (HICP), real GDP growth and unemployment in the euro area for several horizons (see Garcia (2003) for further details). The corresponding real-time observations for inflation and output have been taken from the statistical data warehouse (see Giannone, Henry, Lalik and Modugno (2010) for a detailed description).

In terms of timing, the Eurosystem/ECB forecasts are regularly published early in March, June, September and December. The results from the ECB SPF are regularly published early or in the middle of the months: February, May, August and November.⁵ Given that staff prepares the survey, this gives Eurosystem staff approximately a one month timing advantage (see Bowles, Friz, Genre, Kenny, Meyler and Rautanen (2007)). Figure 2 compares the Eurosystem/ECB's nowcasts with the corresponding euro area SPF nowcasts

⁵See publication schedule on the ECB website: http://www.ecb.europa.eu/stats/prices/indic/forecast/shared/files/SPF_rounds_dates.pdf?cd5f0ee277630018d541b499589fa6f1.

for inflation and output as well as the corresponding forecast errors. Like for the US, this comparison shows that forecasts and corresponding forecast errors are highly correlated for the whole sample since end-2000. This indicates a similar choice of indicators and forecasting strategies. In addition, Table 2 documents that, unlike in the US, the SPF forecasts for euro area inflation are comparable to the Eurosystem/ECB staff forecasts at all horizons considered. But, like for the US, the SPF forecasts for euro area output are inferior to staff forecasts at all horizons considered.

Table 2: Forecast errors for the US and the Euro area

Forecast	MAE	RMSE	ME	Forecast	MAE	RMSE	ME
United States							
Inflation				GDP			
gb qrt0	0.80	1.04	-0.09	gb qrt0	1.34	1.80	0.13
gb qrt1	1.01	1.32	0.07	gb qrt1	1.94	2.66	-0.17
gb qrt2	1.08	1.50	0.12	gb qrt2	2.04	2.85	-0.32
gb qrt3	1.06	1.53	0.08	gb qrt3	1.99	3.00	-0.34
gb qrt4	1.01	1.55	0.03	gb qrt4	1.83	2.75	-0.23
spf qrt0	0.89	1.13	-0.07	spf qrt0	1.50	1.96	0.19
spf qrt1	1.14	1.51	-0.03	spf qrt1	1.85	2.63	-0.12
spf qrt2	1.24	1.69	-0.03	spf qrt2	2.02	2.91	-0.33
spf qrt3	1.37	1.85	-0.04	spf qrt3	2.17	3.19	-0.53
spf qrt4	1.40	1.99	-0.06	spf qrt4	2.06	3.09	-0.61
Euro area							
Inflation				GDP			
ecb yr0	0.14	0.20	0.06	ecb yr0	0.37	0.46	-0.08
ecb yr1	0.64	0.84	0.24	ecb yr1	1.26	1.87	-0.88
ecb yr2	0.68	0.80	0.26	ecb yr2	1.35	2.20	-1.10
spf yr0	0.19	0.26	0.12	spf yr0	0.41	0.55	-0.18
spf yr1	0.59	0.81	0.17	spf yr1	1.34	1.97	-0.99
spf yr2	0.63	0.79	0.09	spf yr2	1.53	2.33	-1.29

Notes: Sample for the Fed: 1968Q4 to 2006Q4; sample for the ECB: 2000Q4 to 2012Q1. MSE: mean square error, RMSE: root mean square error, MFE: mean forecast error, GB: Greenbook forecast, ECB: Eurosystem/ECB forecast, SPF: Survey of Professional Forecasters, qrtn: n quarter ahead forecast, yrn: n year ahead forecast.

4 Econometric strategy and results

In this section we analyze the forecasting performance of staff forecasts and private forecasters for inflation and output, both for each type of forecasters individually and in relative terms. We report empirical results for the US (1968 to 2006) and the euro area (2000 to 2012). First, we test for each type of forecaster whether their forecasts can be considered rational. We report results based on regressions for the full sample and for different sub-samples. Second, we test whether these staff forecasts contain information that is potentially helpful to improve private sector forecasts. Third, we conduct tests for forecast stability of the relative performance of staff forecasts and private forecasts over time. Fourth, we apply tests for conditional predictive ability by which we study alternative exogenous factors that may explain variations in the relative forecasting performance. Given the need for long time series, tests for forecast stability and for conditional predictive ability can only meaningfully applied to US data.

4.1 Forecast rationality

Forecast rationality tests are popular tools when assessing individual forecast performance. These tests check whether a forecast is equal to the mathematical expectation conditional on all (available) information (Muth (1961)). Forecast rationality requires forecasts to be both unbiased and efficient estimates of future outcomes (see e.g. Keane and Runkle (1990)). Since it is not always clear which set of exogenous variables should be included in these tests, it is common to test for weak efficiency, which requires that forecast

errors are not correlated over time. Both, the unbiasedness and the (technical) weak efficiency assumptions have been challenged in the literature. For example, a biased forecast might be optimal in the case of asymmetric loss function (Holden and Peel, 1990), but such forecasts may not be useful seen from the perspective of professional forecasters. Likewise, it has been shown that weak efficiency will falsely be rejected for rational (and efficient) forecasts under a broad range of models. This holds for a data generating process producing frequent persistent level shifts with uncertain timing such as a regime switching error correction model (where the equilibrium level of the dependent variable is observed, but uncertainty prevails about different adjustment regimes). Rational inflation forecasts would then produce correlated errors. Models with this or similar features are frequently applied to a variety of macroeconomic indicators including inflation rendering the validity of the so defined efficiency criterion questionable.⁶ Therefore, while testing for both unbiasedness and weak efficiency, we focus on the results related to unbiasedness, when discussing rationality. We consider two alternative tests to assess unbiasedness of forecasts. First, we employ a standard Mincer-Zarnowitz regression (which is frequently applied, see e.g. Romer and Romer (2000) and Rossi and Sekhpoysan (2011)).

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

where $A_{h,t}$ denotes outcomes at time t (inflation rate or real GDP growth

⁶This argument has been applied to the rationality of inflation forecasts. See for example Johnson (1997), Evans and Lewis (1995) and El-Shagi (2011).

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). The null hypothesis of forecast rationality is $H_0 : \alpha = 0$ and $\beta = 1$ jointly. If the forecasting horizon h is larger than one period ahead, the residuals of this equation will exhibit moving average behavior of an order $h - 1$. This, however, does not indicate inefficiency, because it only reflects that the same shock to the variable of interest shows up repeatedly in the forecast errors since the periods covered by subsequent forecasts overlap. To deal with the arising problem of serial correlation in the forecast errors when estimating equation 1, we calculate robust standard errors for all regressions. Second, we employ a test for unbiasedness as proposed by Holden and Peel (1990). While the condition $\alpha = 0$ and $\beta = 1$ in the test equation 1 is sufficient for unbiasedness, Holden and Peel have shown that it is unnecessarily restrictive. Instead, the condition $\alpha/\mu_A + \beta = 1$ (where μ_A is the mean of A), which is necessary and sufficient for unbiasedness should be used. Whether this condition holds, can implicitly be tested through the equation:

$$A_{h,t} - \hat{F}_{h,t} = \alpha + \eta_{h,t}, \quad (2)$$

where η_t is a moving average process of an order equal to the effective forecasting horizon minus one. Holden and Peel (1990) propose to test the null hypothesis $H_0 : \alpha = 0$. In small samples the Holden and Peel (1990) test could be somewhat problematic. The reduction in degrees of freedom due to the MA terms in the already small subsamples substantially reduces the

power of the test, thereby falsely accepting unbiasedness too often. This problem is aggravated by the fact that dynamic estimation is typically subject to a small sample bias. When comparing the results from the traditional test (equation 1) and the Holden-Peel test it should be borne in mind that the latter test is too restrictive, whereas the former test is too forgiving. We test for weak efficiency by estimating the augmented form of equation 1 (see Keane and Runkle (1990)):

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

where v again follows an MA process of appropriate order. The null hypothesis of weak efficiency is $H_0 : \gamma = 0$, i.e. forecast errors have no predictive power. Under the null, the only dynamic behavior in the forecast errors stems from the MA process and it is caused by the overlapping forecast horizons. In principle, it would be possible to test for unbiasedness ($\alpha = 0$ and $\beta = 1$) and efficiency ($\gamma = 0$) simultaneously in this setup. Like for the Holden-Peel test, an augmented small sample problem may arise due to fewer degrees of freedom after accounting for the AR and MA behavior. Since this would reduce the power of the test, we do not interpret the results obtained from equation 3 as additional check for unbiasedness. When interpreting the results on efficiency, we note that the presence of a small sample bias would imply a bias of the AR term towards zero. Consequently, the test would detect inefficiency less frequently, but it would not falsely detect inefficiencies where these are absent.

Evidence for the United States In terms of forecast rationality, our results for the US using a longer sample of almost 40 years (see Table 3) resemble those obtained earlier on by Romer and Romer (2000). We find that the null hypothesis of forecast unbiasedness cannot be rejected at conventional levels (of 5 percent) neither for the Greenbook nor for the SPF forecasts of inflation and output. That is, both Greenbook and private forecasts contain important information about future inflation and output developments. This finding can be extended to other private forecasters, namely the Blue Chip forecasts (see Rossi and Sekhpoysan (2011)).⁷ We provide additional results on efficiency. For output, the test results support weak efficiency for Fed staff and SPF forecasts. But, for inflation 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 a possible factor which could 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 with longer horizons should not be interpreted as an outright rejection of their rationality.

⁷Due to the fact that for the euro area no Blue Chip data are available we only report tests of the SPF forecasts.

Table 3: Forecast rationality tests (United States)

Horizon (quarters)	α	β	Wald-test	Holden-Peel test	Efficiency test	R^2	N	BP-test
Inflation								
Fed								
0	0.07(0.22)	0.96(0.06)	0.85	0.17	0.51	0.67	152	80Q4
1	0.06(0.29)	1.00(0.08)	0.90	0.65	0.00	0.75	152	80Q4
2	0.08(0.30)	1.01(0.09)	0.80	0.56	0.00	0.68	150	80Q3
3	0.04(0.33)	1.01(0.10)	0.92	0.74	0.00	0.65	144	80Q2
4	0.01(0.38)	1.01(0.12)	0.99	0.86	0.00	0.61	137	80Q1
SPF								
0	-0.30(0.21)	1.05(0.06)	0.30	0.51	0.59	0.82	152	81Q1
1	-0.21(0.31)	1.05(0.09)	0.78	0.91	0.00	0.68	152	80Q4
2	-0.20(0.36)	1.04(0.10)	0.86	0.91	0.00	0.59	152	80Q4
3	-0.16(0.47)	1.03(0.13)	0.94	0.87	0.00	0.51	152	81Q2
4	-0.02(0.55)	0.99(0.14)	0.98	0.89	0.00	0.43	147	81Q1
GDP								
Fed								
0	0.24(0.20)	0.96(0.05)	0.49	0.39	0.78	0.64	152	79Q3
1	0.30(0.34)	0.83(0.10)	0.16	0.53	0.19	0.33	152	79Q2
2	0.10(0.44)	0.85(0.13)	0.20	0.32	0.00	0.19	150	80Q1
3	0.92(0.54)	0.56(0.16)	0.01	0.23	0.71	0.06	144	79Q4
4	0.83(0.61)	0.63(0.19)	0.09	0.38	0.56	0.06	137	92Q4
SPF								
0	-0.11(0.23)	1.12(0.07)	0.10	0.22	0.78	0.64	152	79Q3
1	-0.21(0.39)	1.03(0.12)	0.83	0.65	0.14	0.33	152	79Q2
2	-0.32(0.39)	1.00(0.17)	0.37	0.31	0.00	0.19	152	79Q1
3	0.33(0.81)	0.72(0.22)	0.07	0.15	0.22	0.06	152	79Q4
4	0.05(0.87)	0.79(0.26)	0.04	0.14	0.87	0.06	152	82Q1

Notes: Sample: 1968Q4 to 2006Q4. 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.

Rossi (2005) and Rossi and Sekhpoysan (2011) argue that the above tests for forecast rationality are invalid in the presence of parameter instability.⁸ Changes in paradigms of US monetary policy 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)). 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. 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. In order to check for the existence of a break in the relationship we conduct a break point test.⁹ These tests show that a break has likely occurred at the beginning of the 1980s, i.e. when the Volcker disinflation started (see last column of Table 3). Tests based on rolling window estimation techniques Figure 4 and 5 also indicate occasional or even prolonged departures from the unbiasedness property by Fed staff and by private forecasters. Efficiency is rejected for most periods and for all forecast horizons for inflation and output. However, the above mentioned phenomenon of autoregressive behavior of forecast errors obtained from rational models is particularly severe in small samples as shown by Evans and Lewis (1995), which would render the test invalid.

Evidence for the Euro area Table 4 shows the results for the euro area. The tests tend to confirm forecast rationality both for the Eurosystem/ECB forecasts

⁸We address this issue also in section 4.3 by applying fluctuation tests (see Giacomini and Rossi (2010)).

⁹We used the Andrews, Lee and Ploberger (1996) procedure to test for breaks at unknown time with a trimming parameter of 15%.

and for private forecasters.¹⁰ Rationality of the Eurosystem/ECB staff forecasts cannot be rejected at the 5% significance level, but it is still rejected for the nowcasts (inflation and output) of private forecasters. Given that the regression may suffer from a small sample bias ($n \leq 40$), we apply a dynamic ordinary least squares (DOLS) estimator technique to estimate the parameters.¹¹

Since its creation in 1999, the ECB maintained an unchanged focus on its primary objective of price stability. But, it 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 forecasts alone. We examine this possibility using the rolling window forecasting technique. Figure 6 and 7 show temporary violations of the forecast rationality property for both Eurosystem/ECB staff and private forecasters. Like for the US, we find violations of rationality to be more persistent for inflation than for output forecasts. This evidence shows that the change in the forecast assumption was unlikely the cause for a break in the relationship. Furthermore, visual inspection of the rolling windows estimations shows that during the period of the financial crisis the test properties for rationality of euro area inflation forecasts improved significantly. Such improvements in terms of forecast rationality during the crisis do not mean that forecast errors decreased during the crisis (see Kenny and Morgan (2011)).

¹⁰In the case of the euro area, the results for both the Holden and Peel test and the efficiency test and have to be interpreted with great care. Both tests rely on a moving average structure based on forecasts with a constant horizon. Since current year forecasts, change the forecast horizon every period, an important assumption underlying the test equations could be violated.

¹¹See Stock and Watson (1993) who show that DOLS corrects for small sample bias and leads to unbiased and efficient parameter estimates.

Table 4: Forecast rationality tests (Euro area)

Horizon (years)	α	β	Wald-test	Holden-Peel test	Efficiency test	R^2	N
Inflation							
Eurosystem							
0	0.07(0.05)	0.98(0.03)	0.08	0.11	0.00	0.96	38
1	0.14(0.83)	1.07(0.46)	0.08	0.31	0.00	0.64	36
SPF							
0	0.16(0.14)	0.98(0.07)	0.04	0.01	0.04	0.89	39
1	-0.15(1.02)	1.21(0.54)	0.11	0.44	0.00	0.38	36
GDP							
Eurosystem							
0	-0.06(0.07)	1.03(0.04)	0.71	0.44	0.62	0.95	37
1	-0.88(0.57)	1.10(0.29)	0.11	0.09	0.06	0.43	34
SPF							
0	-0.26(0.10)	1.10(0.07)	0.05	0.07	0.64	0.93	37
1	-0.69(0.77)	0.96(0.37)	0.15	0.06	0.02	0.35	34

Notes: Sample: 2000Q4 to 2012Q1. The standard errors are computed using the Newey-West procedure.

Two explanations for this observation are conceivable. First, noise during the crisis caused an increase in confidence bounds of the point estimates of α and β . This makes it harder to reject rationality. Second, the point estimates moved towards $\alpha = 0$ and $\beta = 1$, because the shocks also implied substantial deviations of macroeconomic indicators from their long run mean. These deviations could in turn be exploited as information for the subsequent forecasts.

4.2 Testing for additional information

Central banks could have information about the economy that is not known to market participants when preparing their inflation and output forecasts. In this section we test whether 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). In order to check for robustness, we also control for a possible timing advantage by the central bank staff. To test for additional information (see e.g. Romer and Romer (2000)), we estimate the following equation:

$$A_{h,t} = \delta + \gamma^P * \hat{F}_{h,t}^P + \gamma^S * \hat{F}_{h,t}^S + \nu_{h,t}, \quad (4)$$

where $A_{h,t}$ denotes outcomes at time t (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. When assessing the possible impact of timing on the relative forecast performance, we repeat the tests from equation 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:

$$A_{h,t} = \delta + \gamma^P * \hat{F}_{h,t}^P + \gamma^S * \hat{F}_{h+1,t-1}^S + \nu_{h,t}, \quad (5)$$

Evidence for the United States Table 5 reports estimation results of equation 4 and 5 for the US. We confirm findings by Romer and Romer (2000) on the potential usefulness of Greenbook forecasts for the private sector. Our results indicate that for an extended sample that 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: Encompassing and timing tests (United States)

Horizon (quarters)	δ	γ^P		γ^S		R^2	N
Encompassing							
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
Timing							
Inflation							
0	-0.29(0.19)	0.99(0.13)	***	0.06(0.13)		0.82	152
1	-0.17(0.26)	0.50(0.18)	***	0.55(0.17)	***	0.69	152
2	-0.14(0.28)	0.25(0.15)		0.80(0.13)	***	0.67	152
3	-0.10(0.30)	0.11(0.14)		0.93(0.12)	***	0.66	152
GDP							
0	0.14(0.25)	1.36(0.11)	***	-0.30(0.12)	**	0.66	152
1	-0.21(0.41)	1.04(0.22)	***	-0.01(0.22)		0.33	152
2	-0.38(0.54)	0.67(0.24)	***	0.35(0.18)	*	0.21	152
3	0.32(0.70)	-0.06(0.24)		0.87(0.12)	***	0.29	152

Notes: Asterisks mark significance at the one(***), five(**) and ten(*) percent level. Since the timing test requires forecasts over a longer horizon for the staff, it cannot be calculated for the longest available forecast horizon of four quarters ahead.

When giving Fed staff a timing disadvantage of one quarter and extending the forecast horizon by one in the test (see equation 5), we find that Greenbook forecasts would no longer provide additional information for short-term forecasts. But, Greenbook forecasts for inflation and output are still encompass the private forecasts at longer forecast horizons, i.e. private forecasts contain no information that is included in Greenbook forecasts. This might indicate that there actually is a deeper understanding behind the structural causes of inflation (and GDP growth) at central banks, because this is required for making good forecasts at longer horizons. Contrarily, the access to the most recent information is essential to perform well in nowcasts.

Evidence for the Euro area Table 6 reports estimation results for euro area inflation and output. These results are rather mixed. Based on forecasts for the current year, the tests indicate that Eurosystem/ECB staff forecasts have additional information not contained in SPF forecasts. However, forecasts for the next year seem not to provide private forecasters with additional information. A further issue here are the small values of R^2 for the next year forecast. This deterioration could be a reflection of the extraordinary uncertainty during the financial crisis (since mid-2007). Moreover, in the case of Eurosystem/ECB forecasts, private forecasters can include the information from the staff forecasts from previous rounds, because they are published and regularly discussed at the monthly press conferences of the ECB.

Table 6: Encompassing and timing tests (Euro area)

Horizon (years)	δ	γ^P	γ^S	R^2	N
Encompassing					
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
Timing					
Inflation					
0	0.24(0.17)	1.09(0.10) ***	-0.15(0.11)	0.90	39
GDP					
0	-0.16(0.10)	1.08(0.08) ***	-0.07(0.07)	0.92	38

Notes: The lead and lag length is thereby determined using the Akaike information criterion. Asterisks mark significance at the one(***), five(**) and ten(*) percent level. Since the timing test requires forecasts over a longer horizon for the staff, it cannot be calculated for the longest available forecast horizon .

When considering the possibility that Eurosystem/ECB staff may have a timing advantage relative to private forecasters, it appears that the above findings may not be robust. In order to account for timing, we lag staff forecasts by one quarter and extend the forecast horizon by one in the test (see equation 5). The results indicate that staff forecasts no longer possess additional information for private forecasters. Since the timing test for the euro area can only be calculated for the current year forecast, it would still be possible that the finding of additional information of staff forecasts is valid for longer forecast horizons. Additionally, it should be considered that current year forecasts comprise of forecasts with relatively short and long horizons. Since the forecast error of nowcasts (such as current year forecasts made in the fourth quarter), is much smaller than the forecast errors over longer horizons, the nowcasts might be the driving force behind the results. Our finding, that central bank staff performs only better at three- and four-quarter-ahead horizons (as in the case of the Fed) is therefore not sufficient to indicate the absence of an informational advantage of the Eurosystem/ECB forecasts.

4.3 Testing for forecast stability

In this section, we examine the stability of relative forecast performance using a recently developed test by Giacomini and Rossi (2010). The test can only be applied to US data in a meaningful way, because it requires sufficiently long time series .

The null hypothesis is forecast stability:

$$H_0 : E[\Delta L_t(\hat{f}_{t-h,R}^S, \hat{f}_{t-h,R}^P)] = 0 \quad \text{for all} \quad t = R + h, \dots, T, \quad (6)$$

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 .¹²

Figure 3 shows the test results for inflation, output and its components for the US. The null hypothesis of forecast stability is rejected, if the test statistics hits one of the confidence bounds shown in this Figure at least once (one-time reversal test). For inflation we find a clear tendency towards instability in the relative performance. The fluctuations of the test statistic reject stability for inflation as measured by the GDP deflator forecasts for all horizons considered. In order to check for robustness, we also make tests for the CPI as proxy for inflation. With the exception of the nowcast, the test also detects instability in the relationship. Since the CPI series provided by the Fed is substantially shorter than the GDP deflator series, the results for the CPI are not fully comparable here. Most importantly, the results for the CPI do not capture the Great Moderation. For (real) output the relative performance was also not stable. When analysing the GDP components, however, we do not find any evidence for instability in the relative forecast performance. The detected difference in behaviour is mostly due to missing observations for GDP components in the early part of our sample. Before the Great Moderation, changes in the relative forecast performance were most pronounced. Although fluctuations in performance do not suffice to reject stability, visual inspection indicates some marked fluctuations for all time series. Prior to the mid-1980s, the fluctuation statistics indicates for most variables superiority of the Fed Greenbook forecasts.¹³ The Greenbook forecasts significantly outperformed the

¹²For details see Giacomini and Rossi (2010) equation (1). As they suggest, we choose the window size to equal 15% of the sample.

¹³Negative (positive) values of the test statistics indicate the superiority (inferiority) of Greenbook forecasts.

private forecasts during the 1970s when the US economy had to face severe oil price shocks. Coinciding with the Great Moderation, which reduced overall volatility, the forecast performance of both types of groups became more similar. The fluctuation test statistics shows that since the mid-1980s there were overall no meaningful differences in the relative forecasting performance for inflation and output.

4.4 Testing for conditional predictive ability

A familiar explanation for superiority of central banks staffs' forecasts is their better knowledge about the future interest rate path.¹⁴ Central bank staff could also have better information about government statistics. For other shocks central bank staff and private forecasters face a similar degree of uncertainty. Examples are oil-price shocks and shocks related to financial crises.

To test for superiority of Greenbook forecasts in the presence of exogenous factors, we use the conditional predictive ability (Wald type) test proposed by Giacomini and White (2006).¹⁵ It provides information on whether changes in the relative forecasting performance are linked to developments in specific exogenous variables. We use a set of conditioning variables Ω_t to proxy four kind of different shocks, namely uncertainty, data revisions, information about the interest rate, and oil¹⁶ and commodity prices.¹⁷ The null hypothesis is that given the information set Ω_t it is not possible to distinguish which forecast group has a lower forecast error at

¹⁴Only few central banks share this information with the public in a systematic manner, e.g. the Swedish Riksbank and Norges Bank. See Rudebusch and Williams (2008) for a detailed analysis.

¹⁵The test requires long runs of data and is therefore only applied to US data.

¹⁶Including oil prices addresses the possibility that a bias in forecasts may have been caused by oil price movements. See e.g. European Central Bank (2012).

¹⁷For the euro area Bowles et al. (2007) address this issue by making a comparison for the HICP excluding and including food and energy.

horizon τ . It can be written as:

$$H_0 : E[L(\hat{f}_{t+\tau}^S) - L(\hat{f}_{t+\tau}^P)|\Omega_t] = 0. \quad (7)$$

First, we examine the relative forecasting performance for inflation and real GDP growth, given an uncertain economic environment. This is proxied by the cross sectional dispersion for the quarterly forecasts (i.e. the dispersion of inflation, real GDP, industrial production, and housing starts).¹⁸ Since the dispersion measures are obtained from the SPF, they could report an information disadvantage that is specific to professional forecasters, because uncertainty perceived by individual forecasters cannot necessarily be deduced from their mutual disagreement. To test for robustness, we also use the predicted variance of inflation obtained from a simple GARCH(1,1) model as a proxy for uncertainty.

Second, we check whether data revisions have an impact on the relative forecasting performance. We use revisions in the variables inflation and real GDP growth.¹⁹ 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. As proxy, we use the absolute quarter-on-quarter change in the federal funds rate at the corresponding forecast horizon. Fourth, we condition on oil prices and the commodity price index.

Table 7 reports the results. First, if we condition the relative forecast performance for uncertainty in the economic environment, we find that, with the notable exception of nowcasts, Fed staff made generally better inflation forecasts than the

¹⁸The dispersion measure equals the 75th percentile minus the 25th percentile of the forecasts for quarter on quarter variables. These variables are available from the Federal Reserve Bank of Philadelphia.

¹⁹This set of revisions is obtained from the real-time data set from the Federal Reserve of Philadelphia.

SPF in times of high uncertainty . While Greenbook forecasts for inflation over longer horizons are found to be more accurate than private forecasts when overall uncertainty is high, its relative forecasting performance for output is only better for nowcasts.

Second, if we condition the relative forecast performance on data revisions, 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 their argument that Fed staff makes better forecasts for reasons which are not related to their earlier access to government statistics.

Table 7: Tests for conditional predictive ability in the United States

Horizon (quarters)	Uncertainty				Data revisions			Int. rate changes	Commodity prices
	DPGDP	DRGDP	DIProd	DHousing	ARCH	RevPGDP	RevRGDP		
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
Real 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: P-values for the conditional predictive ability test are computed following Giacomini and White (2006). P-values below 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, if we condition the relative forecast performance on future changes in the federal funds rate, the results seem to suggest that the Fed made better inflation forecasts during times when interest rate changes were in the pipeline. This result holds for inflation forecasts at a longer horizon of four quarters ahead.²⁰ 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 relative forecast performance 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 central bank and private forecasters.²¹ The test results confirm the widespread notion that both 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

This paper presents new evidence for the US and the euro area on whether central bank staff outperforms private forecasters when forecasting inflation and output. The test results for the US suggest: (i) that Greenbook forecasts used to outperform private forecasts on inflation and output, but (ii) this excess performance has gradually disappeared since the mid-1980s. For the euro area, evidence applies only to the last decade. Our tests suggest that the ECB has an information advantage relative to private forecasters, but the performance of private sector forecasts for inflation and output is not fundamentally different. This assessment inspired us to

²⁰Since 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 the interpretation of our results. 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.

²¹This finding is robust to using different measures for oil. Since the HWWA index also compromises commodity prices we only report the results for this measure in Table 7.

look into the dynamics and the possible reasons for different forecast performance of both central bank staff and private forecasters. We apply a battery of tests to check the rationality of staff forecasts and private forecasts in the US and the euro area. We identify several periods for which forecast rationality disappears for both the Fed and the SPF. During the Great Moderation neither private nor staff forecasts were successful in explaining the variation of inflation around its mean, causing the rationality tests to reject in samples which mostly cover this period. Rationality of the Greenbook inflation forecasts is strongly rejected in many subsamples which include the Volcker disinflation period. This observation is not fully mirrored in tests for private forecasts. It raises the issue why during this episode staff forecasts did not fully exploit all available information. The evidence for the euro area shows improvements in terms of forecast rationality during the financial crisis, when volatility increased dramatically. Partly, this represented noise and caused forecast errors to become large. However, the volatility in the driving forces of inflation and output created new information that could be exploited in forecasting. Rationality tests show that private forecasters in the euro area appear to have had more difficulties with it than Eurosystem staff. Additional tests show that Greenbook forecasts, and less so the Eurosystem/ECB staff forecasts, possess valuable information for private forecasters. While for the US these findings are robust to a possible timing advantage of the Fed, this seems not to hold for the ECB. Unlike for the Fed which publishes its Greenbook forecasts only after five years, in the case of Eurosystem/ECB forecasts, private forecasters can include the information from the staff forecasts from previous rounds, because they are published and regularly discussed at the monthly press conferences of the ECB.

For the US, where sufficient long time series are available, it is possible to examine the reasons for differences in relative forecast performance in more depth. Rel-

ative forecast performance stability tests show that the importance of the Fed's information advantage changed substantially over time. We find that the driving forces behind the narrowing of the information advantage of Greenbook forecasts have coincided with the Great Moderation. Tests for conditional predictive ability show that the differences in the forecasting performance can be explained by factors common to both types of forecasters such as increased economic uncertainty and by specific factors such as the Fed's staff better knowledge of the future interest rate path. These tests also show that other factors such as better knowledge of government statistics and oil price or commodity shocks cannot explain differences in the forecasting performance between Fed staff and private forecasters.

References

- Abreu, I. (2011). International Organisations' vs. private analysts' forecasts: an evaluation, Technical Report 20, Banco de Portugal.
- Alesina, A. and Summers, L. H. (1993). Central Bank Independence and Macroeconomic Performance: Some Comparative Evidence, Journal of Money, Credit and Banking **25**(2): 151–162.
- Andrews, D. W., Lee, I. and Ploberger, W. (1996). Optimal changepoint tests for normal linear regression, Journal of Econometrics **70**(1): 9–38.
- Batchelor, R. (2000). The IMF and OECD versus Consensus Forecasts, Applied Economics **33**(2): 225–235.
- Bikhchandani, S., Hirshleifer, D. and Welch, I. (2008). Information Cascades, The New Palgrave Dictionary of Economics.
- Blattner, T., Catenaro, M., Ehrmann, M., Strauch, R. and Turunen, J. (2008). The Predictability of Monetary Policy, ECB Occasional Paper Series **83**.
- Bowles, C., Friz, R., Genre, V., Kenny, G., Meyler, A. and Rautanen, T. (2007). The ECB survey of professional forecasters (SPF) - A review after eight years' experience, ECB Occasional Paper Series **59**.
- Brand, C., Buncic, D. and Turunen, J. (2010). The Impact of ECB Monetary Policy Decisions and Communication on the Yield Curve, Journal of the European Economic Association **8**(6): 1266–1298.
- Croushore, D. (2012). Forecast bias in two dimensions, Technical Report 12-9, Federal Reserve Bank of Philadelphia.
- 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.
- El-Shagi, M. (2011). Inflation expectations: Does the market beat econometric forecasts?, The North American Journal of Economics and Finance **22**(3): 298–319.
- 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).
- Evans, M. D. and Lewis, K. K. (1995). Do Expected Shifts in Inflation Affect Estimates of the Long-Run Fisher Relation?, Journal of Finance **50**: 225–225.

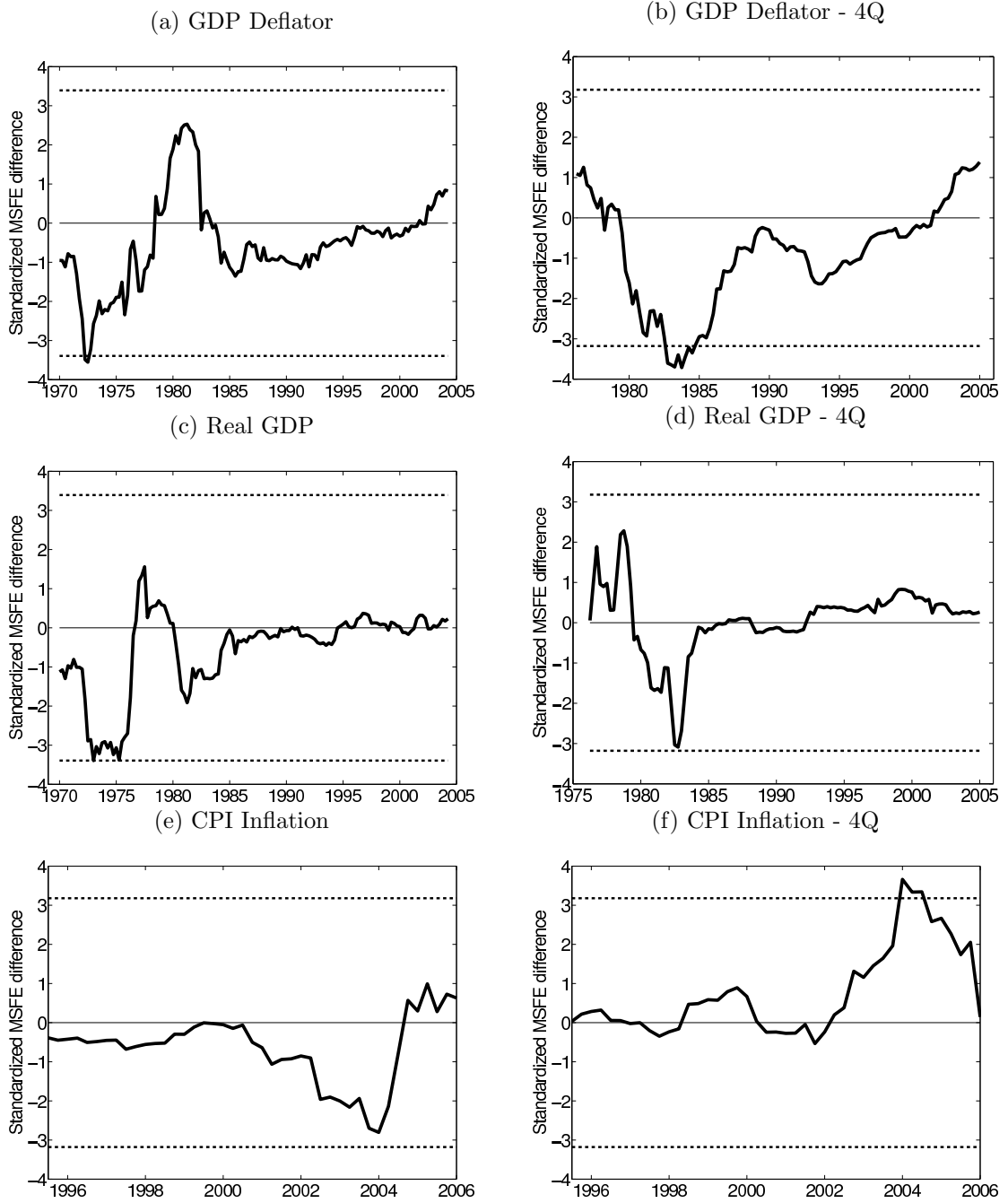
- Frankel, J. and Schreger, J. (2012). Over-optimistic official forecasts in the Eurozone and fiscal rule, Technical Report 18283, NBER.
- Gamber, E. N. and Smith, J. K. (2009). Are the Fed's inflation forecasts still superior to the private sector's?, Journal of Macroeconomics **31**(2): 240–251.
- Garcia, J. A. (2003). An Introduction to the ECB's Survey of Professional Forecasters, Occasional Paper Series 8.
- 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. (2013). Combining expert forecasts: Can anything beat the simple average?, International Journal of Forecasting **29**(1): 108–121.
- 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.
- Holden, K. and Peel, D. A. (1990). On testing for unbiasedness and efficiency of forecasts*, The Manchester School **58**(2): 120–127.
- Hubert, P. (2012). ECB Projections as a Tool for Understanding Policy Decisions, Technical report.
- Johnson, D. R. (1997). Expected Inflation in Canada 1988-1995: An Evaluation of Bank of Canada Credibility and the Effect of Inflation Targets, Canadian Public Policy **23**(3).
- Jung, A. (2011). An International Comparison of Voting by Committees, Working Paper Series 1383, European Central Bank.
- Keane, M. P. and Runkle, D. E. (1990). Testing the rationality of price forecasts: New evidence from panel data, The American Economic Review pp. 714–735.
- Kenny, G. and Morgan, J. (2011). Some lessons from the financial crisis for the economic analysis, Occasional Paper Series 130, European Central Bank.

- Lange, J., Sack, B. P. and Whitesell, W. C. (2003). Anticipations of Monetary Policy in Financial Markets, Journal of Money, Credit and Banking **35**(6): 889–909.
- 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.
- Muth, J. F. (1961). Rational Expectations and the Theory of Price Movements, Econometrica **29**(3): 315–335.
- 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.

- 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, Proceedings, Federal Reserve Bank of Kansas City **Aug**: 399–474.

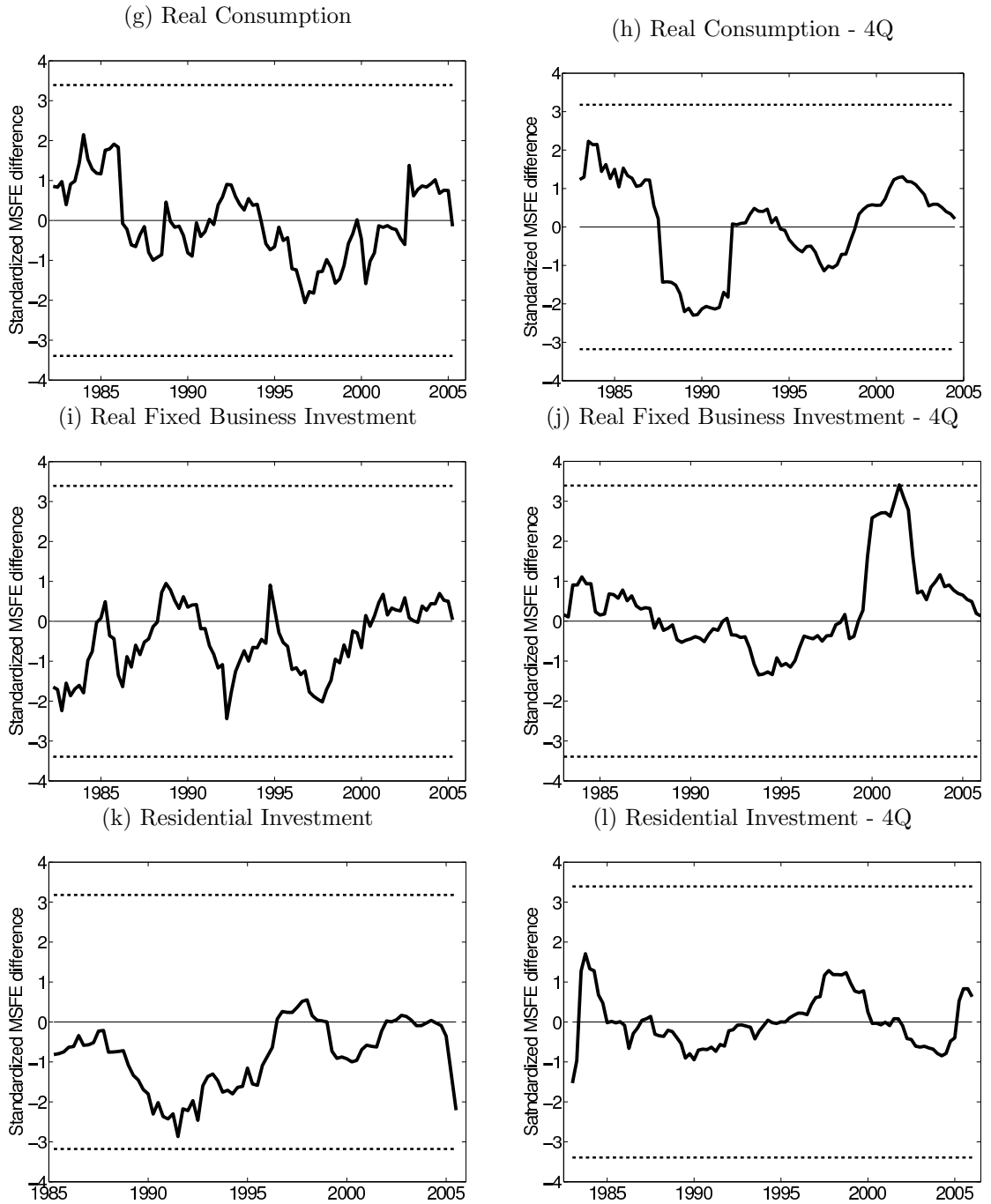
A Figures

Figure 3: Fluctuation test statistic (United States)



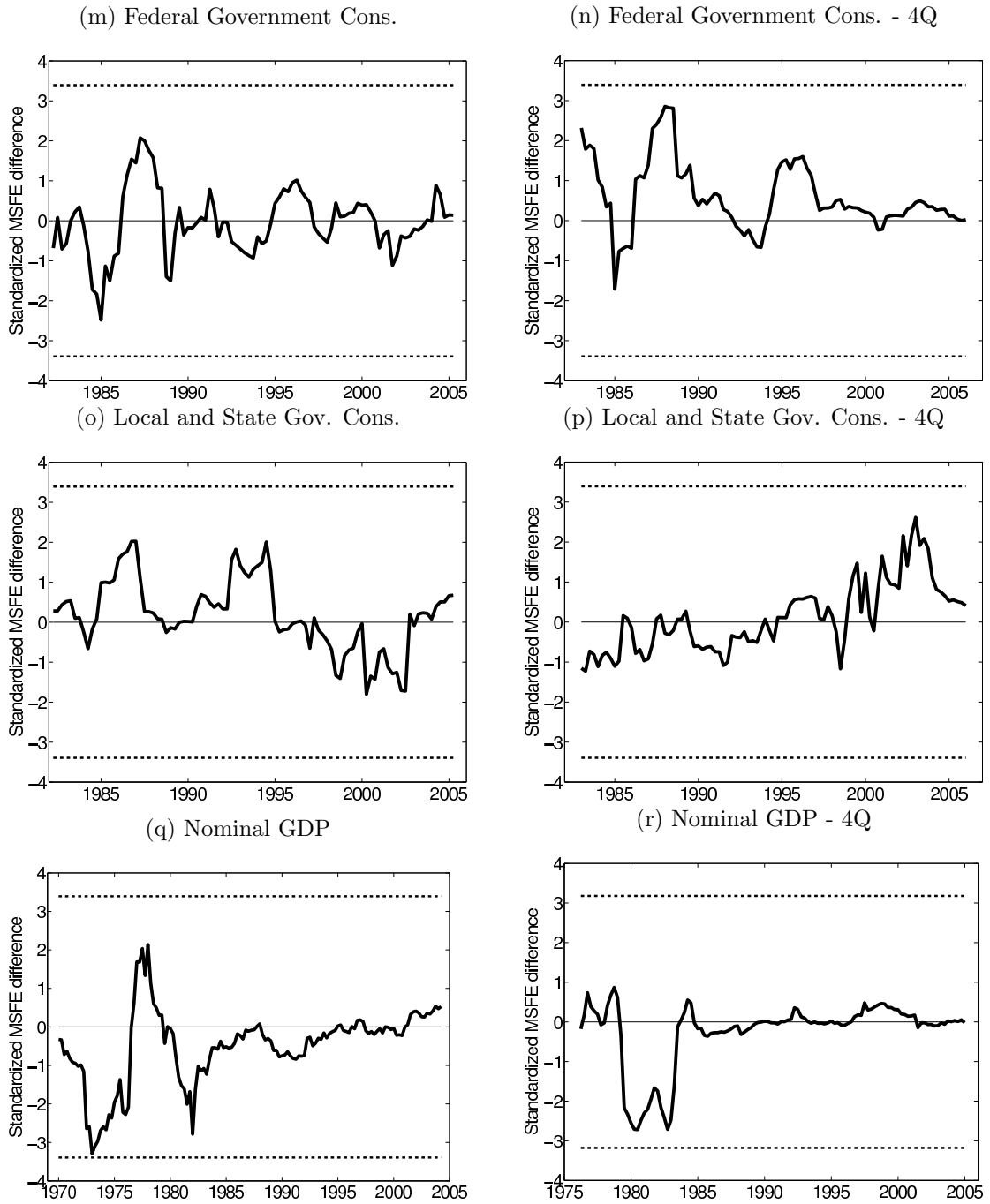
Notes: The solid line shows the fluctuation test statistic and the dashed lines represent the corresponding critical values.

Figure 3 (cont.): Fluctuation test statistic (United States)



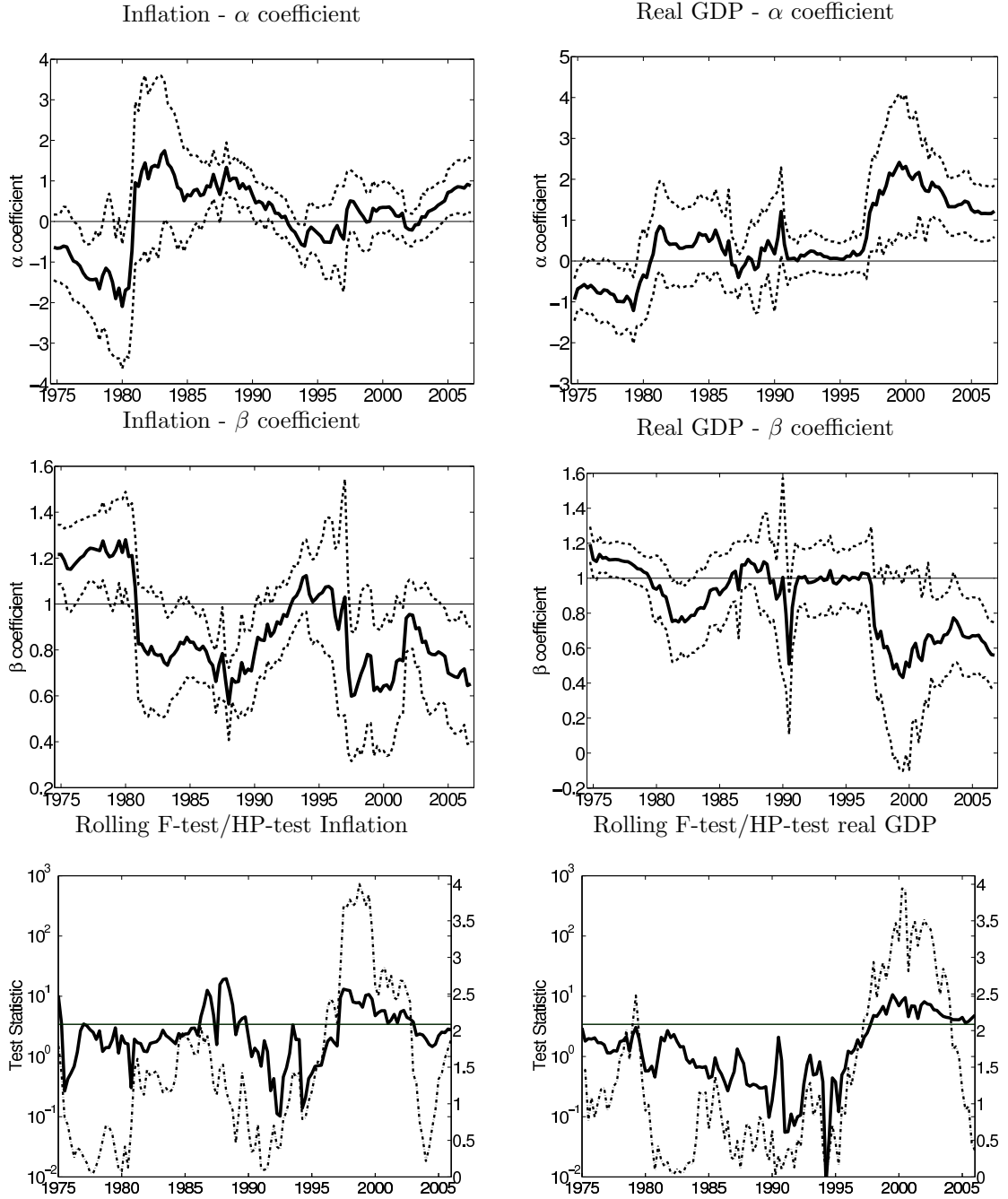
Notes: The solid line shows the fluctuation test statistic and the dashed lines represent the corresponding critical values.

Figure 3 (cont.): Fluctuation test statistic (United States)



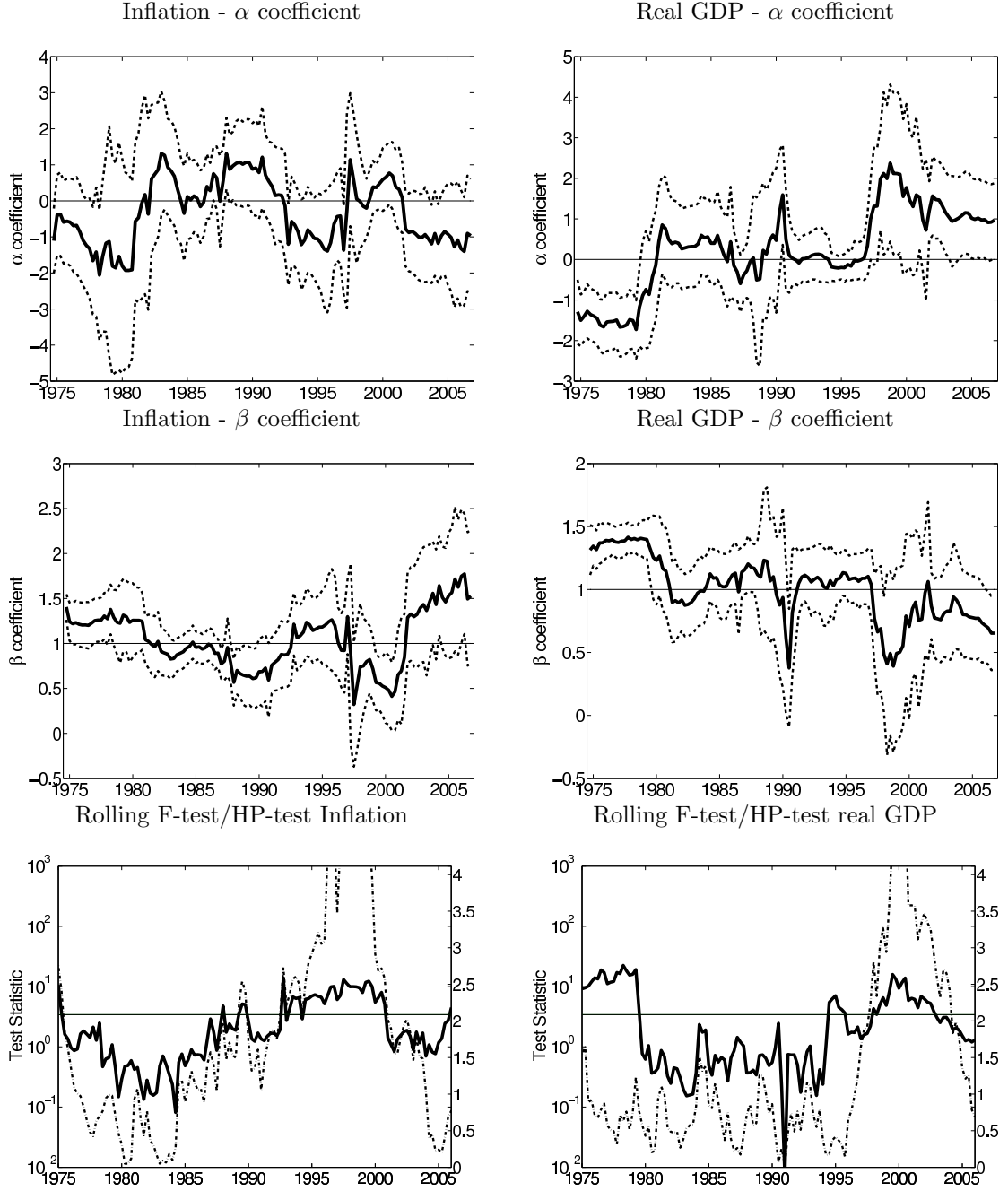
Notes: The solid line shows the fluctuation test statistic and the dashed lines represent the corresponding critical values.

Figure 4: Rolling Window Estimation (Fed Greenbook forecasts)



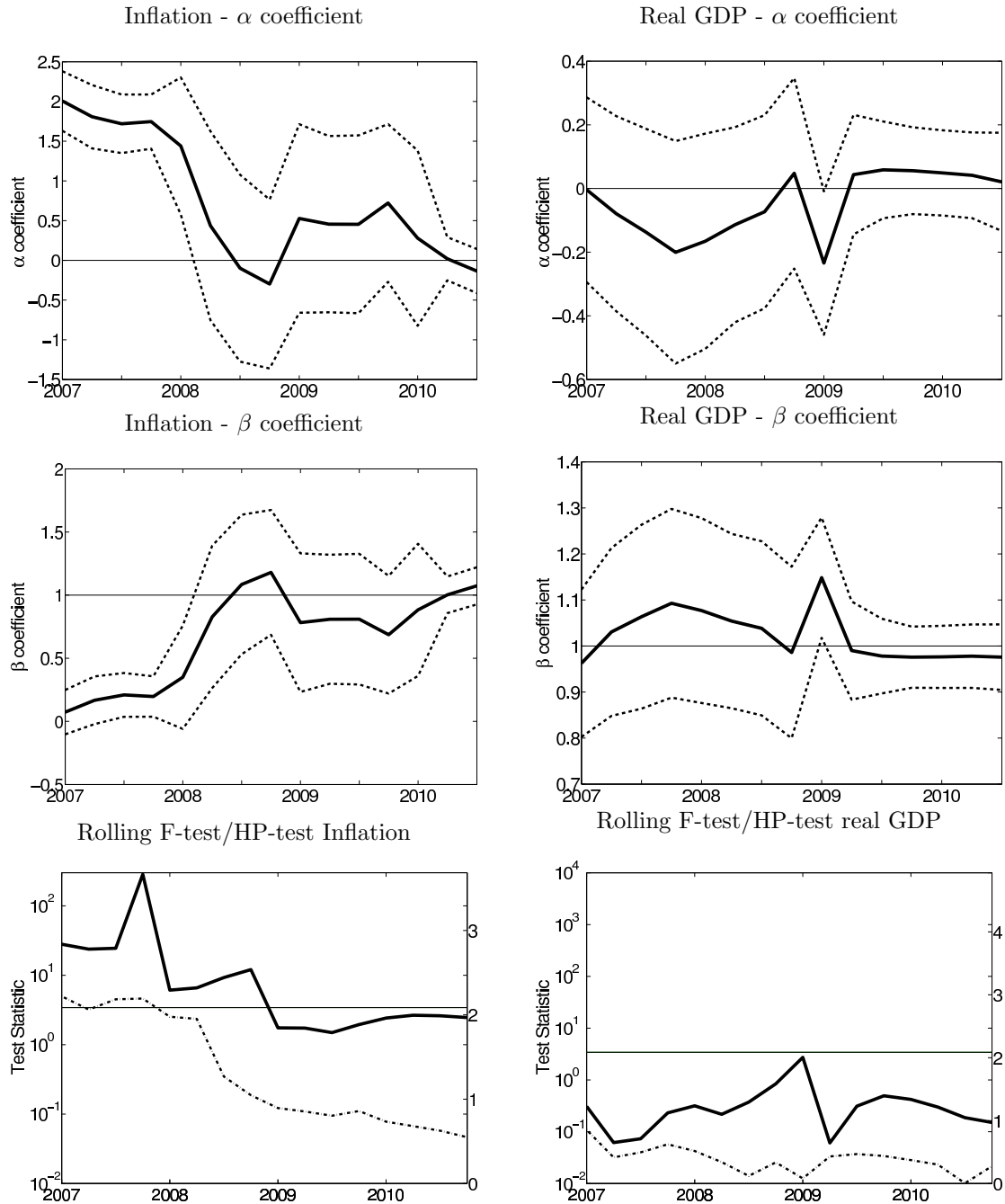
Notes: The top four pictures show rolling window estimates for the individual α and β coefficients (solid lines) with corresponding 95% confidence bounds (dashed lines). The window size for estimation comprises 25 observations. The two bottom pictures show the corresponding evolution of the F-statistic (for the joint hypothesis $\alpha = 0$ and $\beta = 1$; see solid line) and the Holden-Peel-Test (see dotted line). The axes are scaled so that the horizontal line represents the critical value for both tests.

Figure 5: Rolling Window Estimation (US SPF)



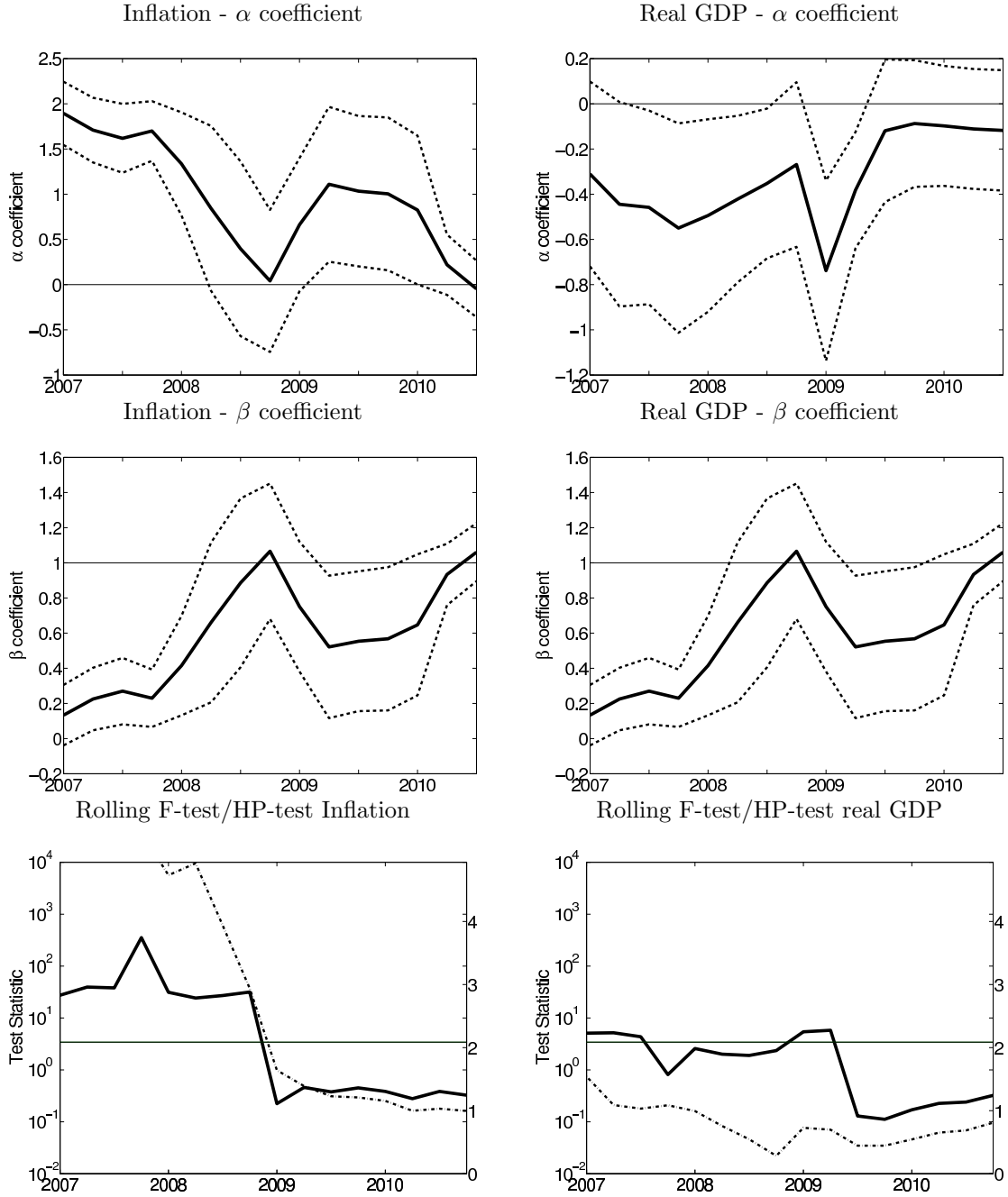
Notes: The top four pictures show rolling window estimates for the individual α and β coefficients (solid lines) with corresponding 95% confidence bounds (dashed lines). The window size for estimation comprises 25 observations. The two bottom pictures show the corresponding evolution of the F-statistic (for the joint hypothesis $\alpha = 0$ and $\beta = 1$; see solid line) and the Holden-Peel-Test (see dotted line). The axes are scaled so that the horizontal line represents the critical value for both tests.

Figure 6: Rolling Window Estimation (Eurosystem/ECB staff forecasts)



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 $\beta = 1$; see solid line) and the Holden-Peel-Test (see dotted line). The axes are scaled so that the horizontal line represents the critical value for both tests.

Figure 7: Rolling Window Estimation (ECB 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 $\beta = 1$; see solid line) and the Holden-Peel-Test (see dotted line). The axes are scaled so that the horizontal line represents the critical value for both tests.