EUROPEAN CENTRAL BANK
EUROSYSTEM

OCCASIONAL PAPER SERIES
NO 59 / APRIL 2007

# THE ECB SURVEY OF PROFESSIONAL FORECASTERS (SPF)

## A REVIEW AFTER EIGHT YEARS' EXPERIENCE

by Carlos Bowles, Roberta Friz, Veronique Genre, Geoff Kenny, Aidan Meyler and Tuomas Rautanen

EUROPEAN CENTRAL BANK

EUROSYSTEM

OCCASIONAL PAPER SERIES

NO 59 / APRIL 2007

# THE ECB SURVEY OF PROFESSIONAL FORECASTERS (SPF)

# A REVIEW AFTER EIGHT YEARS' EXPERIENCE *

by Carlos Bowles, Roberta Friz, Veronique Genre, Geoff Kenny, Aidan Meyler and Tuomas Rautanen


In 2007 all ECB publications feature a motif taken from the €20 banknote.

This paper can be downloaded without charge from http://www.ecb.int or from the Social Science Research Network electronic library at http://ssrn.com/abstract_id=967604.

* *This work was carried out while all the authors were working in the Euro Area Macroeconomic Developments Division of the ECB. The views expressed in this paper are those of the authors and do not necessarily reflect those of the European Central Bank (ECB). We would like to thank Juan Angel Garcia, Hans-Joachim Klöckers, Gerard Korteweg, Bettina Landau, Klaus Masuch, participants at a Eurosystem Monetary Policy Committee seminar and an anonymous ECB referee for their useful comments. Any errors are of course the sole responsibility of the authors. Corresponding author: Aidan Meyler (aidan.meyler@ecb.int). Individual anonymised SPF data are available upon request, for further information please contact ecb-spf@ecb.int*

**Address**
Kaiserstrasse 29
60311 Frankfurt am Main
Germany

**Postal address**
Postfach 16 03 19
60066 Frankfurt am Main
Germany

**Telephone**
+49 69 1344 0

**Website**
http://www.ecb.int

**Fax**
+49 69 1344 6000

**Telex**
411 144 ecb d



ISSN 1607-1484 (print)
ISSN 1725-6534 (online)

# CONTENTS

## ABSTRACT

Eight years have passed since the European Central Bank (ECB) launched its Survey of Professional Forecasters (SPF). The SPF asks a panel of approximately 75 forecasters located in the European Union (EU) for their short- to longer-term expectations for macroeconomic variables such as euro area inflation, growth and unemployment. This paper provides an initial assessment of the information content of this survey. First, we consider shorter-term (i.e., one- and two-year ahead rolling horizon) forecasts. The analysis suggests that, over the sample period, in common with other private and institutional forecasters, the SPF systematically under-forecast inflation but that there is less evidence of such systematic errors for GDP and unemployment forecasts. However, these findings, which generally hold regardless of whether one considers the aggregate SPF panel or individual responses, should be interpreted with caution given the relatively short sample period available for the analysis. Second, we consider SPF respondents' assessment of forecast uncertainty using information from their probability distributions. The results suggest that, particularly at the individual level, SPF respondents do not seem to fully capture the overall level of macroeconomic uncertainty. Moreover, even at the aggregate level, a more sophisticated evaluation of the SPF density forecasts using the probability integral transform largely confirms this assessment. Lastly, we consider longer-term macroeconomic expectations from the SPF, where, as expectations cannot yet be assessed against so few actual realisations, we provide a mainly qualitative assessment. With regard to inflation, the study suggests that the ECB has been successful at anchoring long-term expectations at rates consistent with its primary objective to ensure price stability over the medium term. Long-term GDP expectations – which should provide an indication of the private sector's assessment of potential growth – have declined over the sample period and the balance of risks reported by respondents has generally been skewed to the downside. Consistent with this, respondents have generally reported the balance of risks to their long-term unemployment expectations to be to the upside citing concerns as to whether required structural reforms in the labour market will be undertaken as the main factor behind this risk assessment.

## EXECUTIVE SUMMARY

According to economic theory, expectations play a crucial role in how the economy functions, as well as in how, and the extent to which, policy agents, including central banks, can influence macroeconomic outcomes.[1] At the most fundamental level, this reflects the idea that economic agents, both firms and households, have strong incentives to make their key economic decisions (e.g. consumption, investment or price-setting) in a way that optimally takes into account the likely future developments of the economic variables relevant to such decisions. In line with the importance of expectations in understanding the functioning of the economy, the European Central Bank (ECB), in particular, has a clear interest in being able to assess the private sector's short to medium-term inflation expectations when conducting its regular broad-based analysis and assessment of the risks to price stability. Information about such expectations can, for example, provide evidence on the extent to which shocks affecting the inflation process are perceived by agents as likely to persist or be more short-lived. In addition, it is clearly useful to collect information on private sector forecasts for inflation, as this can serve as a point of comparison with – though no replacement for – the Eurosystem's own assessment of the inflation outlook. Finally, a key ECB concern is also to obtain information that can help assess the extent to which private agents hold expectations that are in line with the ECB's quantitative definition of its price stability objective. Such expectations can help in assessing the credibility of the single currency's monetary policy, as perceived by economic experts located throughout the European Union (EU). To this end, the Survey of Professional Forecasters (SPF) also seeks to gather information about longer-term expectations for the euro area Harmonised Index of Consumer Prices (HICP), the price index on which the ECB announced its performance would be assessed. In order to provide the broader macroeconomic context in which professional forecasters form their inflation expectations, the SPF also collects information on expectations for GDP growth and the unemployment rate over both long and short horizons, as well as probability distributions surrounding the forecast point estimates for all three macroeconomic variables.

This paper constitutes an initial assessment of the information content of the SPF by providing a detailed analysis of the survey data collected over the period between 1999 and 2006. Given that the available sample period is still relatively short, the results of this first comprehensive assessment of SPF data must be treated with some caution. Nonetheless, a number of important insights can be obtained based on the experience over the last eight years.

### SHORT TO MORE MEDIUM-TERM FORECASTING PERFORMANCE

A key result of the analysis presented in this paper is that SPF panel members, like most other forecasters, have tended to underpredict euro area inflation for most of the period since 1999. However, this result warrants careful interpretation. On the one hand, the clear tendency for SPF participants to underestimate the inflation outcome over one and two-year horizons points to the need to exercise a high degree of caution when using such expectations as a predictor of future inflation. On the other hand, our analysis shows that a large part of this "systematic" error can be explained by the sequence of asymmetric and largely unpredictable shocks that hit euro area inflation over the period; oil prices, weather-related food shocks, unforeseen changes in administered prices and indirect taxes. Adjusting for the effects of such shocks, there is far less evidence of a systematic underprediction or bias in the SPF inflation expectations.

We found much less of a one-sided pattern in GDP growth forecast errors. In the early stages

1 Carroll (2001) cites Keynes (1936) as among the first to recognise the importance of expectations within the economic system.

of the survey, respondents underestimated the strength of GDP growth. However, following the slowdown in 2001, GDP growth was then consistently overpredicted by respondents; an error common to many other forecasters. The tendency towards overprediction is most clearly evident for the two-year ahead horizon. Finally, the pattern of errors in the unemployment rate forecasts is somewhat at odds with the errors related to GDP. Respondents have tended to expect a higher unemployment rate than the actual outcome, but if they were to have forecast the real side of the economy according to Okun's Law, one would have expected an error in the opposite direction. However, once account is taken of statistical and methodological revisions in the unemployment rate, the inconsistency between the unemployment rate and GDP growth expectations is substantially reduced. Additionally, the effects of labour market reforms should be taken into account.

More formal regression tests also provide additional evidence of bias in the aggregate SPF results. This evidence is strongest for inflation expectations and weaker for unemployment and GDP growth expectations. Importantly, some of these findings may be subject to a small sample bias. Using a longer sample period, in which the shocks driving inflation might be less asymmetric and one-sided, could reverse this result. Whilst these results confirm the need to exercise caution when considering the information content of the SPF for future macroeconomic outcomes, SPF forecasts generally appear to be slightly superior to naïve and purely backward-looking benchmarks. Lastly, in terms of bias and overall predictive accuracy, the SPF performance over the period analysed is broadly similar to the results of other similar surveys, such as Consensus Economics.

Our analysis of the individual SPF replies suggests that the broad pattern of the individual forecasts is essentially the same as that of the aggregate SPF results. For example, at least 90% of the panel consistently forecast inflation to be below the actual outcome. This may reflect a high degree of commonality in the information available (and not available) to panel members, thus leading them to "get it wrong" not only in the aggregate, but also individually. A similar pattern is observed for both GDP growth and the unemployment rate, in the sense that the overall pattern of forecast errors is broadly and consistently shared among the vast majority of panel members.

Another cross-sectional feature of the SPF panel that we examined was whether there was any perceptible difference between the performance of panel members from the financial sector and the performance of those from the non-financial sector. In general, for all three macroeconomic variables and all horizons, the evidence does not point to any significant differences between these two sectors. Additionally, there is no evidence that economic conditions in the country of the individual panel members have had any significant explanatory power for their euro area expectations. As a final component to our analysis of short to more medium-term forecasting performance, we investigated whether there are forecasters who are systematically better or worse than others. The analysis suggests that a small number of forecasters consistently perform above average for *some* variables and horizons (although not systematically for all variables and all horizons).

## FORECAST UNCERTAINTY

The SPF provides several dimensions for measuring forecast uncertainty. One may rely on information provided by individual point estimates only and consider disagreement among the forecasters' views as an indication of uncertainty, or one may also make use of the probability distributions that each panel member assigns to his/her forecast, which is one of the SPF's most original features. Disagreement among panel members, simply measured by the standard deviation of individual point estimates, does not appear to be a good proxy for overall macroeconomic uncertainty. Across all macroeconomic variables and horizons, the

spread of the cross-sectional distribution is rarely sufficiently wide to encompass the actual outcome. This suggests some limitations to the usefulness of measures of macroeconomic uncertainty that are based on the cross-sectional distribution of point estimates alone, i.e. a high degree of consensus is not necessarily an indication of a low level of forecast uncertainty. For example, over the period 2001-04, the short-term uncertainty about GDP growth, measured as the sum of the average variance of individual distributions and the variance of point estimates, tended to rise and remain high, while the level of disagreement among panel members actually declined.

More generally, since the start of the SPF, the pattern of short-term forecast uncertainty, as perceived by SPF respondents, has been broadly stable for both inflation and the unemployment rate. For GDP growth, however, the SPF has indicated a higher degree of variability, as reflected in a considerable increase in perceived short-term uncertainty over the period 2002-03. For both inflation and the unemployment rate, risks over the one-year horizon have generally been assessed by SPF respondents to be skewed, on average, to the upside over the period 1999-2006. This contrasts with GDP growth for which downside risks have tended to predominate.

There are also some indications that panel members may not fully internalise the overall level of macroeconomic uncertainty. For example, compared with the level of uncertainty indicated by the historical volatility of the actual inflation rate, the perceptions of individual panel members indicate a lower level of uncertainty. In particular, individual panel members tend to give a relatively low weight to outcomes further away from the average. While this possible underestimation of overall inflation uncertainty is much less severe at the aggregate level, our evaluation of the aggregate SPF density forecasts suggests some evidence that respondents failed to accurately assess the risk of higher inflation outcomes over this period. In addition, SPF respondents may have failed to accurately assess the risks of lower growth and higher unemployment associated with the economic downturn. These conclusions suggest a need for caution when using the SPF distributions as an indicator of the overall uncertainty about the macroeconomic outlook. However, it would seem that more definitive and robust conclusions must await the accumulation of further data.

## LONG-TERM EXPECTATIONS AND UNCERTAINTY

Our analysis of the longer-term expectations in the SPF dataset is more descriptive, since such expectations cannot be assessed against actual realisations. There are, nonetheless, a number of findings worth noting. First, a key feature that emerges is the clear anchoring of long-term inflation expectations at a level in line with the definition of price stability adopted by the ECB Governing Council. Throughout each of the eight years analysed, average long-term expectations have remained "below, but close to, 2%", and therefore fully in line with the ECB's quantitative definition of its price stability objective. In particular, average long-term expectations have exhibited considerable resilience notwithstanding the persistently higher-than-expected inflation outcomes over shorter horizons. Rather than revising upward their long-term inflation expectations in response to the numerous short-run inflation shocks, respondents have instead reflected such developments in the balance of risks. In line with these findings, while there has been a significant decline in disagreement about long-term inflation forecasts among SPF participants, uncertainty surrounding each individual forecast has tended to increase slightly over time.

At the level of individual panel members, a number of other important features also emerge from our analysis. First, those individual panel members with expectations towards the lower end of the cross-sectional distribution (around 1.5%) have gradually revised upward their long-term inflation expectations but nevertheless they are still in line with price stability (e.g.

around 1.8/1.9%). This process has resulted in a gradual "narrowing" of the cross-sectional distribution of point estimates, implying higher consensus that outcomes will be below, but close to, 2%. However, another feature of the cross-sectional distribution of long-term inflation expectations was its "bi-modal" nature, giving rise to a relatively high proportion of respondents reporting long-term expectations of 1.8% and 2.0%. Only a small proportion quoted longer-term inflation at 1.9%, although this has changed in 2006. The relatively high proportion of respondents reporting exactly 2.0% or slightly above cautions against any complacency on the part of the ECB in taking its credibility for granted.

Finally, over the longer horizons, our analysis of real variables also reveals some interesting findings. In particular, long-term GDP expectations – which should provide an indication of the private sector's assessment of potential growth – have declined over the sample period. Moreover, the overall level of uncertainty surrounding long-term expectations for GDP growth has been assessed by SPF respondents to have been quite stable over the period 1999-2006, although the balance of risk has generally been skewed to the downside. The uncertainty surrounding the longer-term outlook for the unemployment rate is generally assessed to be considerably higher than for either inflation or GDP growth. In addition, the risks to the longer-term unemployment rate expectations have been persistently assessed by SPF participants to be on the upside. Respondents have linked this to concerns as to whether required structural reforms in the labour market will be undertaken.

# I INTRODUCTION

Around the time of its inception in June 1998, the ECB began preparations – in collaboration with the national central banks (NCBs) in the EU – to establish a survey of the private sector's expectations of future euro area macroeconomic developments, in particular inflation. And thus, the first SPF took place in the first quarter of 1999, immediately following the establishment of the single currency. Since then, up to the fourth quarter of 2006, 32 surveys have been carried out on a quarterly basis, thereby contributing to the growth of a rich source of information on euro area macroeconomic forecasts and expectations.[2] For a more detailed overview of the ECB SPF and its main features, see Garcia (2003), which provided a preliminary assessment of the SPF. In this paper, based on a longer time span of data, we attempt to provide a more comprehensive review of the performance of the SPF to date.

## HISTORY AND MOTIVATION FOR THE SPF

A number of factors underpinned the ECB's wish to initiate the SPF and collect information on the macroeconomic expectations and forecasts of the private sector. In particular, according to mainstream economic theory, expectations play a crucial role in influencing how the economy functions, as well as how and to what extent policy agents, including central banks, can influence macroeconomic outcomes. At the most fundamental level, this reflects the idea that economic agents, both firms and households, have strong incentives to make their key economic decisions (e.g. consumption, investment or price-setting decisions) in a way that optimally takes into account the likely future developments of the economic variables relevant to such decisions.

In line with the importance of expectations in understanding the functioning of the economy, the ECB, in particular, had a clear interest in being able to assess the private sector's short to medium-term inflation outlook when conducting its regular broad-based analysis and assessment of the risks to price stability.[3] Information about such expectations could, for example, provide evidence on the extent to which shocks affecting the inflation process are perceived by agents as likely to persist or to remain more short-lived. Indeed, at the start of Stage III of European Monetary Union (EMU), the whole topic of building forecasts for the euro area macroeconomy represented new ground, with much uncertainty existing about which underlying economic models to adopt and their associated parameters. In this context, it was deemed useful to collect information on private sector forecasts for inflation, as it could serve as a point of comparison with – though no replacement for – the Eurosystem's own assessment of the inflation outlook.

In addition, at the start of Stage III of EMU, a key concern for the new central bank was to obtain information that could help in the assessment of the extent to which private agents held expectations that were in line with the ECB's quantitative definition of its price stability objective. Information on such expectations would help in the assessment of the credibility of the new single currency's monetary policy, as perceived by economic experts located throughout the EU. In this respect, the SPF sought to gather information about medium to long-term expectations for the euro area HICP, the price index on which the ECB announced that its performance would be assessed. At that time, no other survey measures of long-term inflation expectations for the euro area HICP were available.[4] Lastly, in order to

2 In this paper, we use the terms "expectations" and "forecasts" interchangeably.

3 Issing (2003) provides an overview of the two-pillar approach used by the ECB, in the context of the ECB's 2003 review of its monetary policy strategy. See also ECB (2004a).

4 Since then a number of surveys, such as those conducted by Consensus Economics and the Eurozone Barometer, have published information on long-term inflation expectations in the euro area, defined in terms of the HICP. In the United States, the Federal Reserve Bank of Philadelphia has conducted a Survey of Professional Forecasters since 1990 – although this survey actually dates back to 1968 when it was originally known as the ASA/NBER Economic Outlook Survey and was conducted by the American Statistical Association (ASA) and the National Bureau of Economic Research (NBER). See Croushore (1993) for further details.

provide the broader macroeconomic context in which professional forecasters formed their inflation expectations, the SPF was also used to also collect information on expectations for both GDP growth and the unemployment rate at both long and short horizons.

## MAIN FEATURES OF THE SPF PANEL

In interpreting the results of this study, it is important to understand the nature of the SPF data. Table 1 provides an overview of the main features of the SPF dataset. A number of specific features warrant emphasis. First, the survey is currently based on a panel of around 75 forecasters located within the EU.[5] The panel has been established in close collaboration with all the NCBs in the European System of Central Banks (ESCB) in order to ensure that participating forecasters possess the necessary technical expertise. In particular, members of the panel are required to possess macroeconomic expertise relating to the euro area and not just their own national economy. Participants should also have several years' experience in forecasting and in publishing forecasts, preferably for the euro area economy. Lastly, in order to avoid one panel member duplicating another, institutions participating in the panel are required to be independent, i.e. not closely linked to another participant. At present, participants are drawn from the 12 euro area countries plus the three other EU countries (Denmark, Sweden and the United Kingdom) that were EU member states at the time of the surveys' inception.[6] Following the enlargement of the EU in May 2004, the panel was extended in principle to include all 25 EU member countries, although as yet no new participants from these countries have been enrolled.[7]

5 Over time, the composition of the panel has altered and the number of panellists has tended to decline. Some participants have dropped out of the panel, although some have also joined it. Moreover, some mergers between several banks in the panel have reduced the number of participants. All in all, since the beginning of the survey, the panel has decreased from around 90 to around 75 at the current juncture.

6 Quite considerable variation exists in the number of participants across the different EU countries, with Greece having the lowest number of participants (1) and Germany the highest (13). Among the non-euro area Member States, the United Kingdom has the highest number of participants (6) reflecting its pool of experienced forecasters in the financial sector. Importantly, in setting up the panel, although some degree of correlation may be evident, there was no systematic attempt to relate the number of respondents located in each country to the "economic weight" of the different countries in the EU. Rather, participation was mainly determined by the available pool of quality forecasters.

7 At present, there are no panellists from the Member States that joined the EU in May 2004, although the NCBs from these countries were invited in 2004 to nominate new panellists. The main reasons cited for not nominating new panellists were the fact that many financial institutions in the new Member States were linked with existing panellists (and thus would most likely not produce independent forecasts) and that many had not yet built up expertise in forecasting euro area developments.

**Table 1 Overview of main features of the SPF dataset**

| | |
|---|---|
| Frequency of survey | Quarterly survey conducted in January, April, July, October |
| Conducted since | Q1 1999 |
| Variables surveyed | – HICP Inflation (in annual percentage change)<br>– GDP growth (in annual percentage change)<br>– Unemployment rate (in percentage of the labour force) |
| Short to more medium-term expectations | |
| Rolling horizons [1] | one-year ahead<br>two-year ahead |
| Calendar horizons | Current calendar year<br>Next calendar year |
| Long-term expectations | Five years ahead |
| Data requested | Point estimates and probabilities surrounding point estimates (i.e. density forecasts) for all variables and all horizons surveyed |
| Number of currently active panel members | Around 75 |
| Average number of respondents | 59 |
| Geographic domain | EU-wide [2] |

1) The rolling horizons are set one and two years ahead of the latest period for which the variable in question is observed when the survey is conducted and not one or two years ahead of the survey date.
2) As the EU has been enlarged, the survey has been opened to possible experts from the new Member States.

The average number of respondents in each survey round has been 59, although there has been a clear seasonal pattern with, on average, fewer respondents in the third-quarter round, which takes place in July each year. Non-responses may be due to absences or simply to the fact that respondents do not have the time to respond when the survey is conducted. The window between respondents receiving the questionnaire and the deadline for replying is quite short and may not always fit in with respondents' own work schedules and forecast procedures. Nonetheless, the average number of respondents compares quite favourably with other euro area surveys, such as Consensus Forecasts or the Euro Zone Barometer, which have around 30 respondents on average. As an EU-wide survey, the aggregate SPF results therefore reflect the expectations of agents located in countries also outside the euro area.

In terms of its sectoral composition, the SPF panel includes participants from the financial sector (mostly banks), as well as non-financial research institutes and employer or employee organisations. The weight of the different sectors is primarily "supply driven", i.e. it reflects the supply of high-quality experts in the different sectors. In this regard, one feature of the panel is that it is heavily weighted towards the financial sector, with about half of the total panel coming from this sector. In terms of the participants from the non-financial sector, these are mainly from non-financial research institutes, with the representation from social partners organisations (either employer or employee organisations) being more limited (for more details see Garcia, 2003).

In terms of the information collected, as summarised in Table 1, the SPF provides information on expectations for euro area HICP inflation, GDP growth and the unemployment rate over short and more medium to longer-term horizons. Over the short horizon, expectations are collected for both "rolling" and fixed "calendar year" horizons. The rolling horizons are set one and two years ahead of the period (month or quarter) for which the latest official release of a given variable is available. For example, in the fourth-quarter SPF for 2006, which was conducted in October 2006, the latest available HICP data referred to September 2006. Consequently, respondents were asked to report their expectations for the annual rate of change in the euro area HICP in September 2007 and September 2008.[8] As regards the longer-term horizon, participants are asked to provide their expectations for the calendar year five years ahead. For example, longer-term expectations referred to 2011 in the fourth-quarter SPF for 2006.

SPF participants are also asked to assign a probability distribution to their forecasts. This distribution provides information on the probability, expressed as a percentage, of the future outcome being within a specific range. The probability distribution resulting from the aggregation of responses also helps to assess how, on average, survey participants gauge the risk of the actual outcome being above or below the most likely range. Moreover, most of the participants report a set of underlying assumptions regarding the ECB's main refinancing rate, oil price developments and the USD/EUR exchange rate for the next five quarters. Since the second quarter of 2004, participants may also provide quantitative labour cost assumptions for the same calendar-year horizons as for the other main macroeconomic variables. Finally, they can supplement their replies with a qualitative explanation of their survey replies; about a third of the panel members choose to do so on a regular basis.

## FOCUS OF THIS STUDY

In this study, we seek to provide an initial assessment of the data collected in the SPF since its inception in early 1999. Considering

8 Given the difference in the publication lags for euro area GDP, inflation and unemployment, the length of the rolling horizon (when measured from the survey date) varies across variables. Nonetheless, for simplicity, we use the terminology one and two-years ahead when referring to these rolling horizons for all three variables.

the relatively short history of the survey, it is still too early to address definitively many potentially interesting questions, such as whether forecasters are unbiased or whether they incorporate new information efficiently, etc. Our intention is therefore to highlight some of the main features of the information that has been collected over the last eight years. Correspondingly, we analyse the results of the 32 surveys conducted between the first quarter of 1999 and the fourth quarter of 2006 according to the following aspects:

*Short to medium-term forecasting performance:* Section 2 attempts to assess the ability of the surveyed expectations to predict actual future outcomes for inflation, GDP growth and the unemployment rate. Given the significant shocks that have affected the euro area economy over the period 1999-2006 (e.g. successive oil prices increases, substantial exchange rate fluctuations, a global economic slowdown and subsequent recovery), it is of interest to assess the extent to which panel members have been able to accurately predict how such shocks would be transmitted to the economy. We therefore examine and compare the predictive accuracy of the SPF forecasts across different variables and horizons. Given the panel dimension of the SPF dataset, it is also possible to examine the level of heterogeneity in the views of the different SPF participants.[9] The analysis is conducted using standard forecast performance statistics (e.g. to check for any bias in expectations), as well as more sophisticated single equation and panel regressions.

*Forecast uncertainty*: Section 3 of the paper considers the information about macroeconomic uncertainty that can be extracted from the SPF. To capture the main distributional features of the surveyed expectations, different indicators of aggregate uncertainty are considered and compared. We also examine the cross-sectional dimension of the dataset, with a view to assessing the extent to which it can shed light on macroeconomic uncertainty. For example, we assess the extent of disagreement among SPF panel members over time and whether this provides a meaningful indication of forecast uncertainty. Drawing on recent research that has suggested new techniques for evaluating the predictive accuracy of density forecasts, we also provide a preliminary assessment of the extent to which SPF panel members have accurately taken into account the level of macroeconomic uncertainty when forming their expectations.

*Long-term expectations:* The final part of our analysis focuses on the information about longer-term expectations and surrounding probability distributions that are available from the SPF. Such longer-horizon survey expectations do not lend themselves easily to evaluation, not least because the number of observed outcomes against which the forecasts can be assessed is still very low. Therefore, our approach is more descriptive. In particular, we describe the main insights that can be obtained from such indicators. For example, we examine the implications of the evolution of longer-term inflation expectations for the assessment of the credibility of the single monetary policy. In addition, to the extent that they may reflect private sector views about potential growth or the "natural" rate of unemployment, it is also insightful to examine the pattern of longer-term expectations for GDP growth and the unemployment rate.

9 The possible heterogeneity in the expectations of different groups has received growing interest from both policy-makers and researchers. For example, the Swedish national central bank (Sveriges Riksbank) regularly publishes a table in its Inflation Report recording the expectations of seven different groups (money market agents, employer organisations, employee organisations, purchasing managers in trade, purchasing managers in manufacturing, households and firms). For a more research-oriented look into the subject, see Carroll (2001).

## 2 SHORT TO MORE MEDIUM-TERM FORECASTING PERFORMANCE

In this section, we assess the short to medium-term forecasting performance of the SPF to date in three main ways. First, we consider the standard forecast error summary statistics (mean error, root mean squared error, etc). Second, we undertake some statistical tests for unbiasedness. In assessing the results of this analysis, it should be borne in mind that the period under consideration (first quarter of 1999 to the fourth quarter of 2006) is relatively short and one in which there have been a considerable number of shocks to the euro area economy (e.g. oil price shocks, food price shocks, etc). Third, in addition to considering the performance of the aggregate SPF forecasts, we investigate individual SPF forecasts and forecasts for different (national and institutional) sub-groups of the SPF in order to check whether or not the aggregate forecasts are masking significant differences at a more detailed level.

As highlighted in Section 1, SPF respondents are asked to provide forecasts for two types of horizons: rolling horizons and calendar-year horizons. The rolling horizons in the SPF are one year and two years ahead of the latest data available when the survey is carried out. Rolling horizons have two advantages. First, they allow for an assessment of the dynamic pattern of the forecast variables within a particular horizon, which may be hidden by considering only calendar-year forecasts, and second, they provide a larger number of forecasts that can be assessed against the actual outcome than allowed for by considering only calendar-year horizons.[10] However, these rolling horizons cannot be considered "independent" forecasts, as their associated forecast errors partly overlap.[11] Consequently, a shock to the variable being forecast affects the forecast errors for several forecast rounds and not just one. Given the larger number of outcomes for the rolling horizon forecasts, this section focuses on these horizons, although some information is reported on calendar-year horizons as they allow for a comparison with other private sector forecasts.

### 2.1 FORECAST PERFORMANCE STATISTICS FOR THE AGGREGATE SPF

There are a number of standard statistics typically used for evaluating forecasting performance.[12] These statistics are (1) the mean error (ME), (2) the mean absolute error (MAE), (3) the root mean squared error (RMSE), and (4) Theil's U statistic (U). Box 1 provides further information on each of these statistics, as well as a brief discussion on their uses and limitations.

10 An advantage of the calendar-year horizons is that they allow for comparison with other surveys of private sector expectations and forecasts reported by international organisations. Conversely, at present, the number of calendar-year horizons for which data realisations are available is still very small. Since the SPF started in 1999, only eight full calendar years have elapsed.

11 For example, in the first quarter of 2005, SPF forecasters provided a forecast of annual inflation for December 2005, and in the second quarter of 2005, for March 2006. The time periods of the two forecasts overlap, as they both include price changes in the period from March 2005 to December 2005. If some shock were to occur during this overlapping period (say a rise in oil prices), it is likely that inflation would rise sharply, with the result that the forecasts for December 2005 and March 2006 would be too low. For a more conceptual discussion of the problems raised by overlapping data, see Harri and Brorsen (2002).

12 See Timmermann (2006) for a recent and concise review of forecast evaluation criteria and an application to the IMF's World Economic Outlook forecasts.

Box 1

## SUMMARY STATISTICS FOR ASSESSING FORECASTING PERFORMANCE

The mean error (ME) is calculated as the average difference between the forecast value $F_t$ and the actual value ($A_t$) realised.

$$ME = \frac{1}{T}\sum_{t=1}^{T}(A_t - F_t)$$

If the forecasts are unbiased and shocks are symmetric, then the ME should be zero on average. However, if the economy has been subject to a number of one-sided shocks (for example, strong and persistent unexpected increases in oil prices), the ME is unlikely to be close to zero over short time periods. A positive value for the ME indicates that, on average, over the whole run of forecasts, the actual value of a particular variable was underestimated, i.e. the forecasts were too low. A negative ME, on the other hand, indicates, that the forecasts were too high on average. One disadvantage of the ME statistic is that it averages out positive and negative errors. Thus, it is possible that two forecasters could have the same mean error, but that one could make larger positive and negative errors on average. As a result, the ME statistic only provides limited information on the overall accuracy of the forecast.

The mean absolute error (MAE) statistic overcomes this shortcoming by calculating the average of the absolute errors. Thus, the MAE statistic gives information on the average size of forecast errors, whether they are positive or negative.

$$MAE = \frac{1}{T}\sum_{t=1}^{T}\left|A_t - F_t\right|$$

The MAE statistic is not without shortcomings either. Consider, in particular, the viewpoint of a policy-maker using forecasts as a decision guide. It is possible, for example, if there are non-linearities, that one or two forecasts that are very far from the actual outcomes could be more damaging from a policy viewpoint than a larger number of forecasts that are only a small distance from the actual outcomes. If this is the case, the policy-maker would like an error statistic that penalises outliers. The root mean squared error (RMSE) statistic, which squares errors before averaging them, thus provides an alternative measure of forecast accuracy. Considering two forecasters with the same average absolute error, the RMSE statistic more heavily penalises the forecaster who makes some large errors compared with the forecaster who makes mainly small errors.

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(A_t - F_t)^2}$$

An additional forecast statistic that may be considered is Theil's U statistic, which compares the RMSE error of a forecast against a benchmark forecast of a random walk – what we will refer to as a "naïve" forecast (which takes the "current" value of the forecast variable).[1] The idea behind this statistic is that although a forecast may look good relative to other forecasts, it could be that simply assuming that the forecast is equal to the last available value could produce a better forecast. Indeed, this is quite often the case for variables such as exchange rates and oil prices.

1 In this note, Theil's U statistic is calculated as the ratio of the RMSE of the aggregate SPF forecast to the RMSE of the "naïve" alternative forecast. A value greater than one means that the "naïve" forecast performs better.

**Table 2 SPF forecast errors (Q1 1999 to Q4 2006 survey rounds): summary statistics**

(annual percentage change unless otherwise indicated)

| | Inflation | | Real GDP | | Unemployment [1] | |
|---|---|---|---|---|---|---|
| **Sample statistics** | | | | | | |
| Actual value | | | | | | |
| Mean (1999-2006 [2]) | 2.1 | | 2.0 | | 8.4 | |
| Standard deviation (in p.p.) | 0.5 | | 1.2 | | 0.5 | |
| **Forecast error statistics** | **1-year ahead** | **2-years ahead** | **1-year ahead** | **2-years ahead** | **1-year ahead** | **2-years ahead** |
| Forecast value | | | | | | |
| Mean | 1.7 | 1.8 | 2.1 | 2.4 | 8.7 | 8.4 |
| Standard deviation (in p.p.) | 0.2 | 0.1 | 0.6 | 0.3 | 0.7 | 0.6 |
| ME (in p.p.) | 0.5 | 0.4 | -0.1 | -0.8 | -0.5 | -0.1 |
| MAE (in p.p.) | 0.5 | 0.5 | 0.9 | 1.1 | 0.6 | 0.7 |
| RMSE (in p.p.) | 0.6 | 0.5 | 1.0 | 1.3 | 0.8 | 0.9 |
| Theil's U | 0.9 | 0.7 | 0.7 | 0.8 | 1.4 | 1.1 |

Sources: ECB, Eurostat and ECB calculations.
1) As a percentage of the labour force.
2) Q3 2006

Table 2 reports the main forecast error summary statistics for inflation, real GDP growth and the unemployment rate, looking at the rolling horizons (one-year and two-year ahead).

Considering first the results for inflation, the ME statistic indicates that inflation has tended to be underestimated over both the one and two-year ahead horizons. This is not so surprising, as HICP inflation since 1999 has been affected by a number of upward shocks (oil price increases in 1999-2000 and 2004-06, unprocessed food price shocks linked to BSE and foot-and-mouth disease in 2001, unexpected increases in administered prices and indirect taxes). The one-year ahead rolling annual inflation forecast underestimated inflation by 0.5 percentage point on average, as indicated by the ME. The underestimation was fairly consistent over time (i.e. there were not many offsetting positive and negative forecast errors) since the MAE is broadly similar at 0.5 percentage point. The RMSE, at 0.6 percentage point is slightly larger, indicating some, albeit not much, variability in the size of errors across survey rounds. More strikingly, although one may expect RMSEs to increase along forecast horizons, given that it is more difficult to make accurate forecasts over longer horizons, the RMSE for two-year ahead inflation forecasts is slightly smaller than that for one year ahead. Overall, the results confirm the particular difficulty that SPF forecasters encountered in foreseeing the cluster of temporary shocks to inflation since 1999.

In contrast to the inflation forecasts, both one and two-year ahead real GDP growth rates tended to be overestimated by an average of 0.1 percentage point and 0.8 percentage point respectively according to the ME. The lower one-year ahead ME statistic conceals fairly large errors, both positive and negative, since the MAE equals 0.9 percentage point, as well as some limited variability in the size of the errors, as the RMSE is only slightly larger than the MAE.

At first glance, the average overestimation of the unemployment rate (0.5 percentage point for the one-year ahead horizon) is at odds with the overestimation of GDP growth. Based on Okun's law, a negative relationship would normally be expected between GDP growth and unemployment. However, the large overestimation of the unemployment rate is mainly an artefact from downward revisions to the unemployment data, which were particularly large in 2000 and 2001.[13]

13 One should also consider the impact of structural reforms in the labour market.

These revisions mostly reflect statistical changes that could not have been predicted. Using the data available at the time the forecasts were realised, the average root mean square forecast error is reduced substantially from 0.5 percentage point to 0.1 percentage point. The influence of statistical and methodological changes on forecast performance highlights one of the pitfalls in forecast evaluation, particularly when there are only a small number of realised forecasts. However, the impact of data revision is much less significant for inflation and GDP growth forecasts.

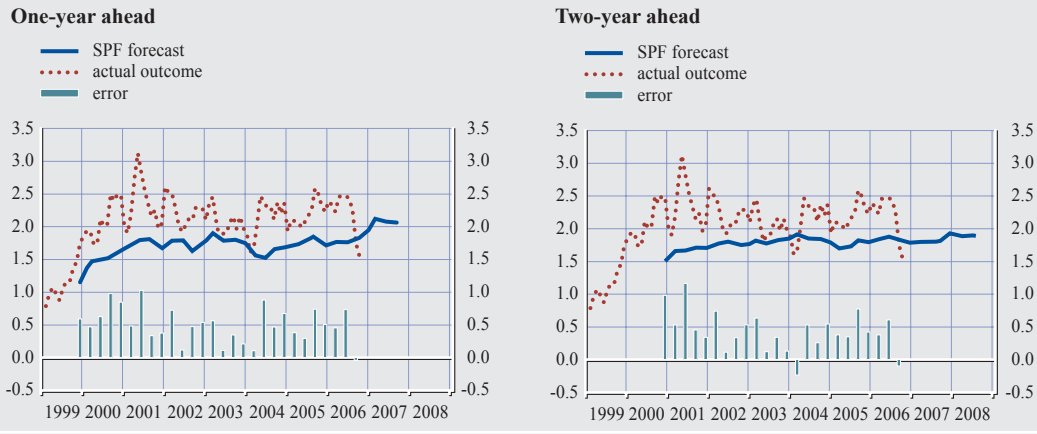
On average, one-year ahead forecasts are more accurate than two-year ahead forecasts, apart from in the case of inflation, where they appear to be more or less comparable. Table 2 also reports an SPF Theil's U statistic of below 1 for both the one and two-year inflation and real GDP growth horizons. This suggests that SPF forecasts for inflation and GDP growth are more accurate than a naïve forecast and that SPF replies contain information about the future, beyond what is already contained in the most recent data.[14]

Chart 1 plots the detailed forecast errors for inflation, which have been clearly and persistently positive over the sample period, with the exception of the observation for the fourth quarter of 2006. In interpreting this result, it should be recalled that the period since 1999 has been characterised by a cluster of large (and generally one-sided) shocks to inflation. The most significant of these have been oil and non-oil commodity price developments, large exchange rate movements, food price shocks associated with animal health concerns and weather conditions, the euro cash changeover and significant changes to administered prices and indirect taxes. It is therefore insightful to consider each of these factors and their correlation with the forecast errors in more detail.

Inflation forecast errors, in particular, appear to be correlated with oil price movements. For example, oil price increases in 1999-2000, reinforced by exchange rate movements, contributed to a rise in headline inflation that translated into substantial forecast errors. The forecast errors were particularly large in the second half of 2000, illustrating the persistence of errors over several survey rounds. Similarly, in 2001, overall inflation was heavily influenced by movements in volatile price components, in particular unprocessed food prices, driven up by health concerns related to BSE and the

14 We tested this hypothesis further using Granger causality tests, which are a more rigorous test of predictive content, but the results did not confirm evidence that survey forecasts had predictive content for future developments. Such a result may be due to the small sample size, however.

**Chart 1 One and two-year ahead inflation forecasts**




Sources: ECB, Eurostat and ECB calculations.

outbreak of foot-and-mouth disease in a number of euro area countries. The measures taken to prevent the diseases spreading further (culls, export bans, etc.) exerted further upward pressures on meat prices, which contributed significantly to an unanticipated pick-up in headline inflation. Consequently, forecast errors in 2001 were also relatively high compared with the average. In the first half of 2004, SPF forecast errors were once again quite large. This reflected increases in administered prices (e.g. related to the German health care reform) that were decided upon and announced in late 2003. All in all, it seems that the persistent underestimation of euro area inflation by SPF forecasters is very much related to the series of unanticipated upward shocks to inflation that has occurred since 1999. As different shocks impacted at different times, it is difficult to disentangle the different factors graphically. Nonetheless, the impact of oil price developments, food prices and the impact of government measures on inflation forecast errors over the one-year horizon is clear.[15] As many of these shocks were difficult to predict, this analysis puts into context the observed persistent underestimation of inflation.

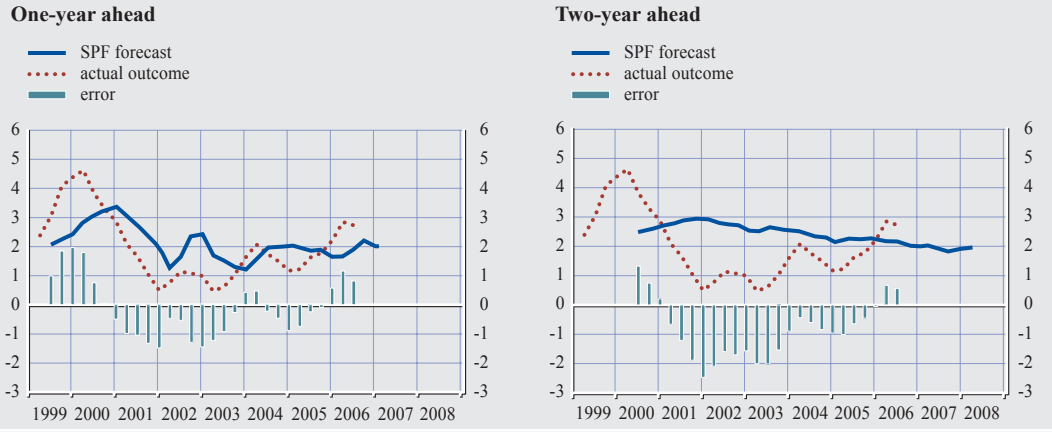
Chart 2 below plots the series for one and two-year ahead GDP growth forecast errors over the sample period.[16] With the forecast errors being mainly negative since early 2001, GDP growth expectations have also been somewhat one-sided, although less so than inflation.

In this regard, GDP growth may have been less asymmetrically affected by shocks during the period under consideration. Particularly at the beginning of the period, GDP appears to have been subject to some favourable shocks. SPF forecasters underestimated output growth in 2000. This may have been due to the positive impact of the exchange rate depreciation on exports combined with the fact that the adverse effects of 1998's emerging economy crisis on the global economy appeared to have been more limited and short-lived than expected. For most of the period since 2001, however, SPF forecasters have overestimated GDP growth.

15 This assessment is also confirmed by an econometric investigation that sought to link the forecast errors with (1) the annual rate of change in oil prices, (2) the annual rate of change in unprocessed food prices and (3) the impact of government measures on HICP inflation. In each case, a positive and significant relationship was found with the forecast error.

16 Compared with inflation, data for GDP growth are much more affected by publication lags. Hence, whilst the GDP forecasts are referred to as the one and two-year ahead forecasts, it is important to note that they, in fact, refer to four and eight quarters ahead of the latest data observation for GDP growth. Given the data publication lags involved, this means that forecasters actually have as much as five months worth of relevant information (such as industrial production or survey data) and this may give some insight into GDP developments. Hence, in reality, the forecasts for GDP growth may be more like two and six-quarter ahead forecasts. When comparing forecast performance across variables, it is important to keep these differences in mind.

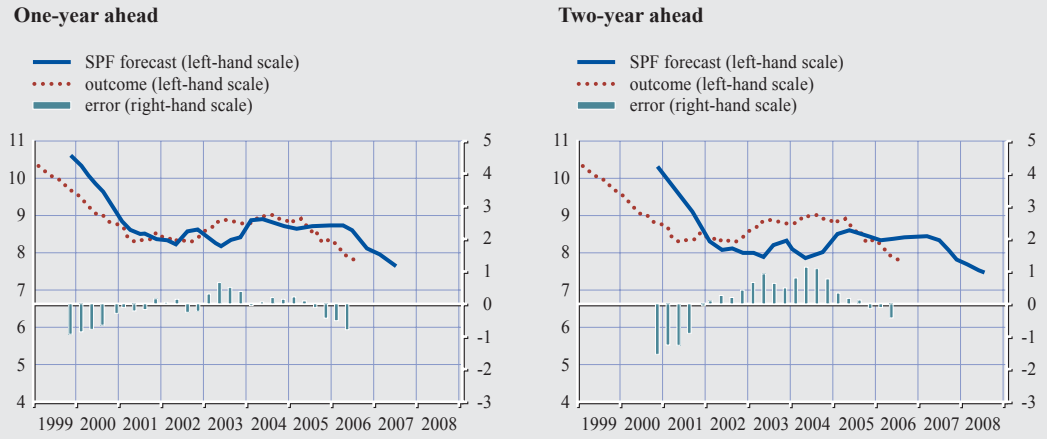
**Chart 2 One and two-year ahead forecasts of real GDP growth**




Sources: ECB, Eurostat and ECB calculations.

**Chart 3 One and two-year ahead forecasts of the unemployment rate**

(using real-time data[1])




Sources: ECB, Eurostat and ECB calculations.
1) To control for the impact of data revisions, we show in the chart the first vintage (or release) of unemployment data. For example, unemployment data released in January 2002 for November 2001 reported an unemployment rate of 8.5%. However, the current estimate of unemployment in November 2001 is 7.9%, approximately 0.6 percentage point below the initial estimate. These revisions may be due to changes in methodology or additional more complete data becoming available subsequent to the initial estimate.

This might be explained by the clustering of unexpected adverse shocks which heightened uncertainty and reduced the magnitude of output growth. In 2000 and 2001, the protracted increase in oil prices impacted negatively on consumption and investment. Moreover, the deterioration in the external environment of the euro area, stemming mainly from the United States, curtailed growth of euro area exports and investment. At the end of 2001, the terrorist attacks in the United States on 11 September were a further blow to business and consumer confidence. In addition, the sharp fall in equity prices, the continued decline in confidence and the geo-political tensions in the run-up to the Iraq war contributed to a further two years of disappointing real GDP growth performance.

Finally, at first glance, unemployment also appears to have been strongly overestimated in 2000 (by around 1.5 percentage points), but this is largely an artefact from downward revisions to the unemployment rate. As these revisions mostly reflect statistical changes that could not have been predicted, Chart 3 reports the forecast errors using unrevised data (i.e. first estimates). The average forecast error is then substantially reduced, especially for one-year ahead forecasts. The profile of the unemployment forecasts mirrors the downward movement observed in the unemployment rate between 1997 and 2001, which reflected a combination of cyclical and structural factors (such as an increasing incidence of part-time work, structural reforms in the labour market, etc.) and the broad levelling-off thereafter.[17]

## 2.2 TESTING FOR BIAS IN THE SPF

We can further assess the accuracy of SPF predictions by presenting statistical tests for the unbiasedness of the forecasts. Unbiasedness implies that forecast errors are zero on average.[18]

17 For a recent overview of longer-term perspectives on structural unemployment in the euro area, see ECB (2005b).

18 As a rule, if forecasts are in line with the Rational Expectation Hypothesis (REH) formulated by Muth (1961), they should be unbiased. The REH states that market participants use all cost-efficient knowledge to forecast economic variables so that their forecasts will tend to be unbiased and efficient. Efficiency requires that the forecasts use all relevant information available at the time that they were made. Tests for efficiency include checking the forecast errors for (lack of) serial correlation and orthogonality tests (whether the forecast errors are independent of the information set used in making the forecast). An additional and related property of rational forecasts is consistency (i.e. each forecast reflects only new information available since the last forecast). However, due to the overlapping nature of the forecasts, neither efficiency nor consistency are tested here as the sample is still too small to obtain robust results.

**Table 3 Regression tests for bias in the SPF aggregate forecasts**

| Variable | α | β | $R^2$ | F-test (p-value) | μ (p-value) |
|---|---|---|---|---|---|
| Inflation one year ahead | 1.06 (0.38) [1] | 0.68 (0.22) | 0.16 | 0.000 | 0.000 |
| Inflation two years ahead | 3.74 (0.92) [1] | -0.86 (0.53) [1] | 0.09 | 0.000 | 0.000 |
| GDP growth one year ahead | -0.33 (0.59) | 1.09 (0.34) | 0.29 | 0.722 | 0.646 |
| GDP growth two years ahead | 4.10 (1.64) [1] | -0.98 (0.65) [1] | 0.08 | 0.001 | 0.005 |
| Unemployment one year ahead | 7.79 (1.05) [1] | 0.07 (0.11) | 0.01 | 0.000 | 0.065 |
| Unemployment two years ahead | 11.82 (0.78) [1] | -0.40 (0.09) [1] | 0.45 | 0.000 | 0.824 |

Source: ECB calculations.
Note: Calculated using actual outcomes during the period 1999-2006. Standard errors are in brackets (corrected using the Newey-West procedure). The p-value is from the F-test; coloured figures indicates rejection of unbiasedness at the 10% level or lower.
1) Indicates that the coefficient is significantly different from the null hypothesis (i.e. α = 0 or β = 1) at the 10% level or lower.

The standard test of unbiasedness is to test the joint hypothesis α = 0 and β = 1 in (1) using an F-statistic:

$$A_t = \alpha + \beta F_t + \varepsilon_t \qquad (1)$$

where $F_t$ is the forecast and $A_t$ the actual realised value. If we cannot reject the null hypothesis, we can conclude that the forecast is unbiased. Holden and Peel (1990) proposed a further test for unbiasedness that is performed directly on the forecast errors and tests whether μ is equal to 0 in (2):[19]

$$A_t - F_t = \mu + \varepsilon_t \qquad (2)$$

Table 3 reports the results of tests applied to equations (1) and (2) for SPF aggregate forecasts. The probability values of the F-statistic for most variables tested are extremely low and unbiasedness is generally rejected (i.e. evidence of bias is found). However, it should be noted that the point estimates for $\alpha$ and $\beta$ are difficult to interpret. Hence, equation (2) may be a better test. The last column of Table 3 reports the probability values for the t-statistic testing whether μ is equal to 0. We still find strong evidence that SPF inflation forecasts are biased. This is not surprising considering inflation forecasts have always been significantly below actual inflation since the beginning of the survey. Looking at GDP growth, it seems that only two-year ahead forecasts are biased, while growth forecasts for one-year ahead appear to be unbiased. Finally, turning to unemployment, we can reject the joint hypothesis α = 0 and β = 1 for all forecast horizons. However, we cannot reject that the mean of the forecast error for unemployment two years ahead is zero. Overall, the statistical evidence is quite weak and not fully robust owing to the small number of observations.

It is useful to amend equations (1) and (2) in order to investigate whether the results obtained are sensitive to the cluster of unanticipated exogenous shocks that would generate large outliers in the observed forecast errors. A formal way to investigate and confirm this assumption would have been to use dummies for each identified shock, but this would have considerably reduced an already very small sample. To circumvent the issue, we ran bias tests, replacing overall HICP inflation by HICP excluding energy and unprocessed food, and run the same regressions as in Table 3. Although SPF forecasters never intended to forecast HICP excluding energy and unprocessed food, this artifice can be seen as an alternative to shock dummies to the extent that many of the shocks between 1999 and 2006 affected energy and unprocessed food prices. Table 4 provides the results of the bias tests from these regressions.

19 As forecasts are made for horizons that are longer than the frequency of forecast rounds (i.e. one-year ahead forecasts vs quarterly rounds), the estimated residuals ($\varepsilon_t$) are serially correlated. As a result, the standard errors of the estimated coefficients will be underestimated and the t-statistics, overestimated. They therefore need to be corrected for any inference to be drawn from the F-tests. We used the Newey-West variance estimator to perform this correction.

**Table 4 Regression tests for bias in the SPF aggregate inflation forecasts**

(Dependent variable: HICP excluding unprocessed food and energy prices)

| Variable | α | β | $R^2$ | F-test (p-value) | μ (p-value) |
|---|---|---|---|---|---|
| Inflation one year ahead | -0.72 (1.04) | 1.50 (0.61) | 0.22 | 0.492 | 0.413 |
| Inflation two years ahead | 0.94 (1.62) | 0.56 (0.92) | 0.02 | 0.401 | 0.212 |

Source: ECB calculations.
Note: Calculated using actual outcomes during the period 1999-2006. Standard errors are in brackets (corrected using the Newey-West procedure). The p-value is from the F-test.

As suspected, although the estimated coefficients differ from the theoretical values predicted by the assumptions of unbiasedness, we find that there would be less of a bias if the HICP excluding energy and unprocessed food prices is used. This suggests the unprocessed food and energy price shocks to inflation explain a large part of the downward bias visible in SPF replies.[20]

The above tests suggest evidence of a statistically significant bias in the aggregate SPF results. However, testing for bias in the aggregate results does not take account of the individual micro dimension to the SPF dataset. In particular, it has been shown that running tests at the aggregate level can introduce a potentially severe aggregation bias in the results (see for example, Keane and Runkle, 1990). Annex 1 therefore considers the results of tests for bias using the full panel of SPF data, thereby enhancing the test by taking into account the full set of cross-sectional information in the dataset. The analysis shows that using a panel of individual forecasts reveals an even more significant bias than the aggregate results would suggest. Consequently, these results generally highlight the widespread difficulty that even professional forecasters had in making accurate forecasts under recent macroeconomic conditions.

## 2.3 COMPARING SPF AND CONSENSUS ECONOMICS FORECASTS

Given the short period of time for which SPF forecast errors are available, it is difficult to assess to what extent the observed bias is due to poor forecasting or the occurrence of persistently unpredictable events over the sample period. In this context, it is also useful to consider the relative forecasting performance of the SPF compared with other forecasters. Such a comparison can reveal whether or not the members of the SPF were poor at forecasting per se or whether their errors were shared with others. If the latter is the case, this could suggest that specific shocks or structural changes that were difficult to foresee were behind the SPF forecast errors rather than a poor absolute performance on the part of SPF panel members.

The SPF forecasts can be assessed against a comparable benchmark, i.e. the averages of private sector forecasts published by Consensus Economics.[21] We consider the results from the SPF and Consensus Economics January surveys every year and compare forecasts for the current and the following calendar year. Two small

20 We also checked whether data revisions may have impacted the findings by using real-time data (i.e. initially released data) at each point, instead of what the latest data vintages indicate. This did not change the results significantly, although it slightly weakened the finding of unbiased growth forecasts one year ahead and suggested that unemployment rate forecasts were unbiased when compared with the initially released data.

21 Every month since 1989, Consensus Economics has published forecasts for major economic variables prepared by a panel of 10 to 30 private sector forecasters, initially for the G7 countries and subsequently for over 70 other economies. The forecasts are published in the second week of each month, based on a survey of panellists' forecasts in the previous two weeks. Together with the individual forecasts for each variable, Consensus Economics also publishes their arithmetic average, the so-called "consensus forecast", for that variable. These predictions relate to the current year and the following year. They can thus be directly compared with the calendar year aggregate forecasts provided with the SPF results, using January forecasts for the current year (t) and the following calendar year (t+1) from both datasets. In interpreting the comparison between the SPF and Consensus Economics, it should be borne in mind that the two surveys share some common panel members.

**Table 5 Forecast errors (SPF, Consensus Economics)**

(January surveys)

| | Current calendar year (t) | | Following calendar year (t+1) | |
|---|---|---|---|---|
| | Inflation | | | |
| | Mean error | RMSE | Mean error | RMSE |
| SPF | 0.3 | 0.3 | 0.5 | 0.5 |
| Consensus Economics | 0.3 | 0.4 | 0.5 | 0.5 |
| *Naïve forecast* | *0.2* | *0.4* | *0.3* | *0.7* |
| | Real GDP growth | | | |
| | Mean error | RMSE | Mean error | RMSE |
| SPF | -0.1 | 0.7 | -0.7 | 1.3 |
| Consensus Economics | -0.1 | 0.7 | -0.8 | 1.4 |
| *Naïve forecast* | *-0.2* | *1.0* | *-0.4* | *1.5* |

Sources: Consensus Economics and ECB calculations.

differences need to be highlighted. First, Consensus Economics has only been reporting euro area figures since January 2003. Prior to this date, Consensus forecasts correspond to the aggregation of euro area country forecasts. Second, unlike the SPF, Consensus forecasts may reflect Consumer Prices Indices (CPI) and not specifically HICPs. Table 5 shows statistics on average mean errors and RMSEs for HICP inflation and real GDP growth for the current calendar year and the following calendar year. It is important to note that there are only a small number of outcomes against which to compare forecasts (seven for the current calendar year forecast and six for the following calendar year forecast).[22] Overall, Table 5 shows that both Consensus Economics and the SPF provide comparable forecasts in terms of accuracy.[23]

Looking at the current calendar year inflation forecast and the following calendar year GDP growth forecast, the SPF appears to perform marginally better than Consensus Economics. However, this is due primarily to rounding differences. On the other hand, although the error statistics for the following calendar year inflation forecast appears the same, in fact the Consensus Economic forecasts perform marginally better than the SPF to the second decimal point. In any case, such small differences in forecast performance when making comparisons over such a short period mean that the differences between Consensus Economics and SPF results are not statistically significant. Moreover, RMSE results show that both surveys slightly outperform a naïve forecast. Overall, it is clear from this comparison that participants in the SPF did no worse than many other forecasters over the period.

## 2.4 HETEROGENEITY AND INDIVIDUAL FORECAST PERFORMANCE

The analysis above focused on the aggregate SPF results. These results can conceal considerable heterogeneity among the respondents. In this section, we focus on the forecast heterogeneity and, in particular, on whether taking into account this heterogeneity can provide additional insight beyond that provided by the aggregate numbers.

### SUMMARISING INDIVIDUAL FORECASTS

Given the large number of respondents to the SPF, summarising the individual responses in a manageable way is not straightforward. Box-whisker charts are one useful way of depicting and summarising the heterogeneity among individual respondents in the context of a survey such as the SPF. Each box-whisker plot contains summary information about the individual replies, as illustrated in Box 2.

22 Results for the unemployment rate are not reported, as prior to January 2003 Consensus Economics only reported forecasts for the three largest euro area economies.

23 The same exercise could have been carried out using Euro Zone Barometer survey results, but the euro area forecasts have only been available since 2002.

Box 2

## INTERPRETING THE BOX-WHISKER CHART

Both the mean and the median of the SPF point estimates are presented in the box-whisker chart. The median is the point forecast with exactly half of the panel members above and half below. Differences between the mean and the median of the point estimates might indicate the presence of outliers in one direction, as well as potential future changes. For example, in response to a shock or new information, it could be that, initially, only a number of forecasters change their point forecasts, which could, although not necessarily, have a bigger impact on the mean than on the median. If all other forecasters were to incorporate this information in the subsequent rounds, this would result in changes in the aggregate forecast that might be signalled by movements in the mean relative to the median.

In addition, the "inter-quartile" range of forecasts is indicated by the extremities of the box. The inter-quartile range depicts the range between the 25th and the 75th percentile in the distribution of point estimates. This is often considered as a useful summary measure of the dispersion of point estimates as it is not affected so much by outliers or extreme values.

The "whiskers" represent the 10th and 90th percentile of the distribution of the point estimates. These provide additional information on the spread of point estimates and, hence, on the level of disagreement among panel members. In particular, the size of the whiskers relative to the box can provide some insight into the share of forecasters with expectations that are some distance away from the "middle ground". Lastly, where available, we plot the actual data outcome for comparison purposes.

**Illustration of the information provided by the box-whisker chart**

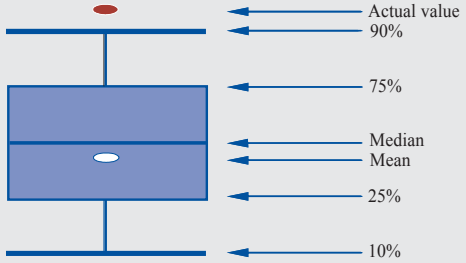
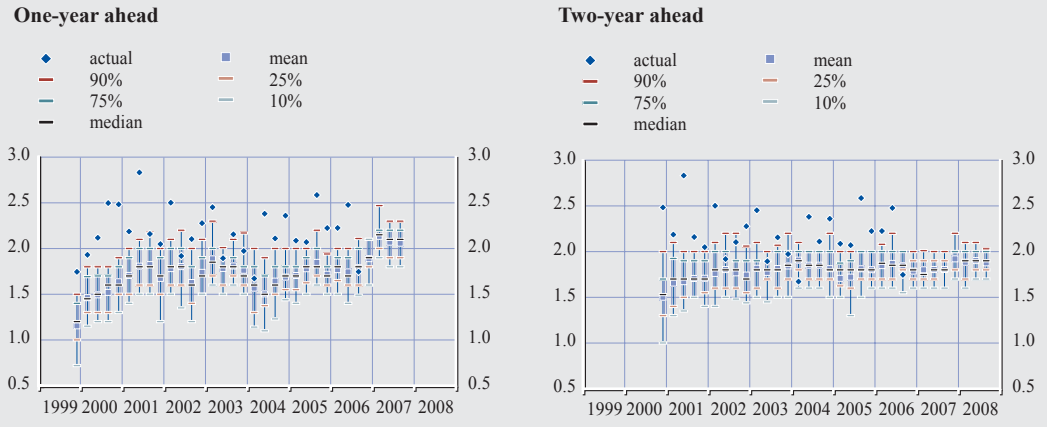



Chart 4 presents box-whisker plots for one and two-year ahead inflation forecasts. The broad pattern of the individual forecasts is essentially the same as that of the average SPF results. Considering first of all the one-year ahead forecast, it is clear that the vast majority of SPF panellists underestimated inflation. Even though the range between the 90th and 10th percentiles was 0.63 percentage point on average (the average inter-quartile range was 0.33 percentage point), for the period under consideration, the actual outcome lay outside even the 90th percentile for much of the time (22 out of 28 realised values lay outside the 90th percentile). Only in the fourth quarter of 2006 was the actual outcome below the median (50th percentile). Thus, although the individual SPF forecasts exhibited substantial heterogeneity, the actual outcome was systematically above the majority of panellists' forecasts.

In the macroeconomic literature, a positive correlation between the level and the variance of inflation has frequently been found;[24] this

24 See for example, Golob (1994) for US data, Crawford and Kasumovich (1996) for Canadian data or Kontonikas (2004) for UK data. Note, however, that these studies generally consider the relationship between inflation and inflation uncertainty over long periods of time (from the 1970s onwards) and can thus cover different inflation regimes. Our sample is limited to a period of relatively low and stable inflation.

Chart 4 **Box-whiskers plots of inflation forecasts (rolling horizons)**




Sources: ECB, Eurostat and ECB calculations.

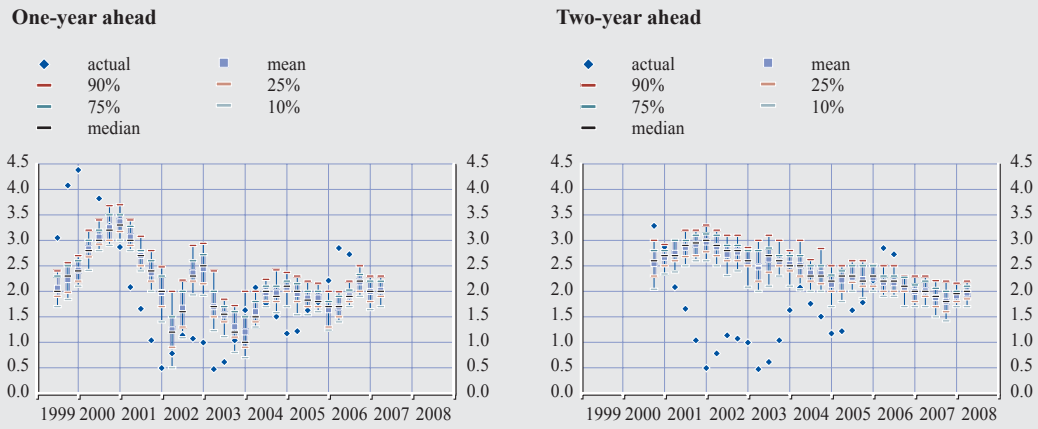
could lead to the *a priori* expectation that there is a positive relationship between the dispersion of inflation forecasts and the mean forecast.[25] However, the correlation between the mean of the forecasts and the dispersion of the forecasts (as measured either by the inter-quartile range or the range between the 90th and 10th percentiles) has been negative. This may reflect the impact of common shocks observed over the period; for example, an increase in oil prices should affect individual forecasts in the same direction, thus increasing the mean but not necessarily the dispersion of forecasts.

Turning to the two-year ahead forecasts, once again inflation was invariably underestimated by a large majority of forecasters. The inter-quartile range (at 0.29 percentage point) and the range between the 90th and 10th percentiles (at 0.55 percentage point) turn out to be slightly lower than those for the one-year ahead forecasts. At first glance, this might appear somewhat surprising, as one might expect the range of estimates to increase over longer horizons. However, this lower level of dispersion and variability in the two-year ahead forecasts suggests that they are relatively anchored, in line with the view that monetary policy can play a role in stabilising inflation over this horizon. Furthermore, the dispersion of point estimates once again does not appear to be positively correlated with the mean of the point estimates.

Turning to the GDP forecasts, the average value of the inter-quartile range for one-year ahead GDP forecasts was 0.40 percentage point, which is somewhat larger than that for inflation over the same horizon. This indicates a larger dispersion of GDP growth forecasts. Similarly, the range between the 90th and 10th percentiles was 0.79 percentage point on average, also above the range recorded for inflation. The respective ranges for two-year ahead GDP forecasts were 0.35 percentage point and 0.71 percentage point, which are both smaller than the one-year ahead forecast. Whilst the actual value was frequently outside the range of the 90th to 10th percentiles for one-year ahead forecasts this was, unlike for inflation, both on the upside and the downside. By contrast, over the period for which data are available, the actual outcome has tended to be below the 10th percentile for the two-year ahead GDP forecasts, with the exception of the first and more recently realised outcomes. Lastly, and again similar to the inflation forecasts, the correlation between both the inter-quartile and

25 See for example, Kiley (2000) and D'Amico and Orphanides (2006).

**Chart 5 Box-whisker plots for real GDP growth forecasts (rolling horizons)**




Sources: ECB, Eurostat and ECB calculations.

90th to 10th percentile ranges and the mean was negative for one-year ahead GDP. It was positive, but close to, zero for the two-year ahead GDP growth forecasts. It is also worth noting for the two-year ahead forecasts that the mean of GDP forecasts has been, on average, below the median, which implies that the distribution of point estimates was skewed to the downside (see Chart 5).

Lastly, turning to the unemployment forecasts, unlike the inflation and GDP forecasts, both the inter-quartile and the 90th to 10th percentile ranges increase as the forecast horizon increases (see Chart 6). The former was 0.30 percentage point on average for one-year ahead forecasts and 0.43 percentage point for two-year ahead forecasts. The corresponding values for the 90th to 10th percentile ranges were 0.57 percentage point and 0.87 percentage point. The fact that the ranges increase as the forecast horizon increases most likely reflects the "non-stationary" nature of the unemployment rate. In other words, both inflation and GDP growth tend to be relatively "stationary" in that they tend to return to a certain range as they are anchored by monetary policy and the rate of growth of potential output, respectively.[26] On the other hand, the unemployment rate may move either up or down on a more permanent basis, as has been observed over the past four decades.[27] There was also a small positive correlation between the ranges and the mean forecast, unlike for inflation and GDP. In particular, the downward movement in the unemployment rate forecast over the period was associated with a growing consensus on the unemployment outlook.

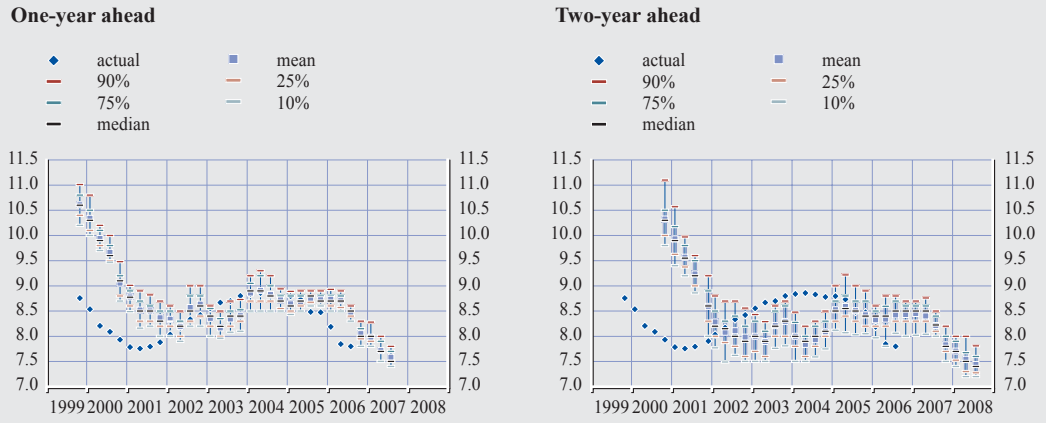
## INDIVIDUAL FORECAST PERFORMANCE

It may also be interesting to consider the heterogeneity of forecast performance in terms of the individual RMSEs of panel members. The results reported in Table 6 show considerable variation in terms of forecast accuracy. For example, in the case of inflation one year ahead, the worst forecaster has an RMSE that is more than twice that of the best forecaster. A similar difference across forecasters is observed for other variables and horizons, with the spread between the best and worst forecaster being smallest for one-year ahead GDP. However, some caution is warranted when ranking forecasters based on such a small number of

26 Both the ECB and the European Commission have communicated extensively on their estimates of potential growth, which may have helped to anchor SPF respondents' longer-term growth expectations. See, for example, ECB (2005a) and European Commission (2006).

27 Whilst these movements, in the first instance, reflect cyclical factors, structural features of the labour market determine the persistence of these movements (see Blanchard 1997).

**Chart 6 Box-whisker plots for unemployment forecasts (rolling horizons)**




Sources: ECB, Eurostat and ECB calculations.

observations. For example, the "best" ranked forecaster for two-year ahead GDP (who nevertheless had a considerable RMSE of 1.1 percentage point) was the most pessimistic, and although *ex post* it seems that such pessimism may have been warranted, this may not have been the case ex ante, given the unforeseen shocks that occurred during the period. Similarly, the "best" forecaster of the one-year ahead inflation rate has consistently forecast above average inflation (even when actual inflation was close to or below the average) and has missed some of the turning points in inflation.[28] Overall, one should stress that there may be other dimensions for assessing the individual performance of SPF forecasters that cannot be captured by the RMSE, such as the frequency with which forecasters reply, whether they provide information on either the assumptions or the probability distribution underlying their forecast, or whether they provide some qualitative insight into their forecasts (see Annex 2 for a more detailed discussion of alternative indicators for assessing the "quality" of SPF forecasters).

Overall, given the short sample for analysing individual forecasters' performance and characteristics, the results should be interpreted with caution. Even though a small number of forecasters show above-average performance across a number of the indicators presented, the sample period is too short to provide a robust test.

28 In addition, some of the "better" inflation forecasters did not report inflation forecasts for the period when inflation was significantly higher than expected. The fact that forecasters do not generally reply to all rounds and for all forecast variables and horizons complicates such a comparative analysis.

**Table 6 Heterogeneity of forecast performance (rolling horizons)**

(root mean squared errors)

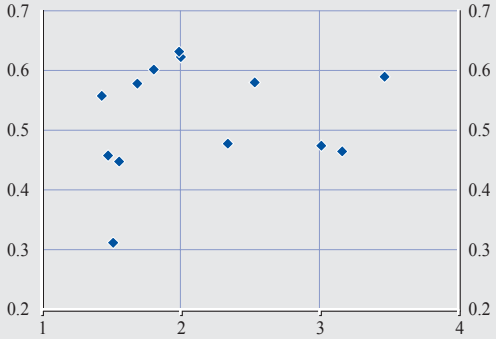
| | Inflation rate | | Real GDP growth | | Unemployment rate | |
|---|---|---|---|---|---|---|
| | 1-year ahead | 2-years ahead | 1-year ahead | 2-years ahead | 1-year ahead | 2-years ahead |
| Minimum | 0.4 | 0.3 | 0.9 | 1.1 | 0.4 | 0.5 |
| Average | 0.6 | 0.5 | 1.0 | 1.3 | 0.8 | 0.9 |
| Maximum | 0.9 | 0.8 | 1.3 | 1.7 | 1.1 | 1.3 |

Source: ECB calculations.

**Chart 7 The relationship between average forecast errors and national developments**

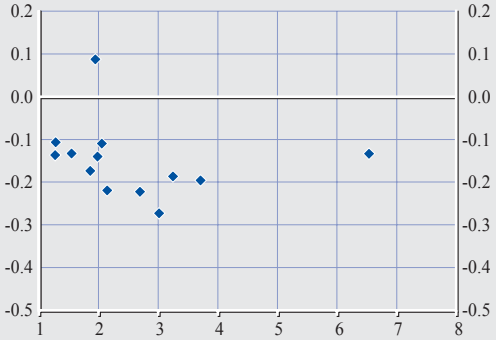
**Correlation between SPF respondents' forecast error and average national inflation rate**

x-axis: average inflation 1999-2006*
y-axis: mean error one-year ahead (in percentage point)



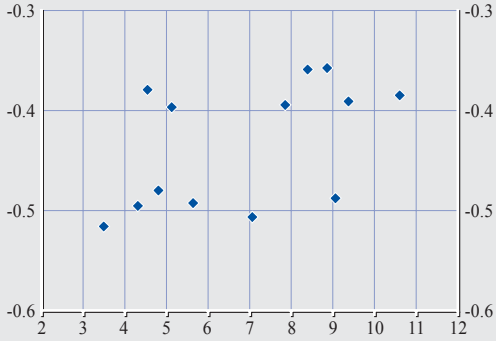

**Correlation between SPF respondents' forecast error and average national real GDP growth**

x-axis: average GDP growth 1999-2006*
y-axis: mean error one-year ahead (in percentage point)




**Correlation between SPF respondents' forecast error and national unemployment level**

x-axis: average unemployment rate 1999-2006*
y-axis: mean error one-year ahead (in percentage point)




Sources: Eurostat and ECB calculations.

## FORECAST PERFORMANCE AND NATIONAL LOCATION OF PANEL MEMBERS

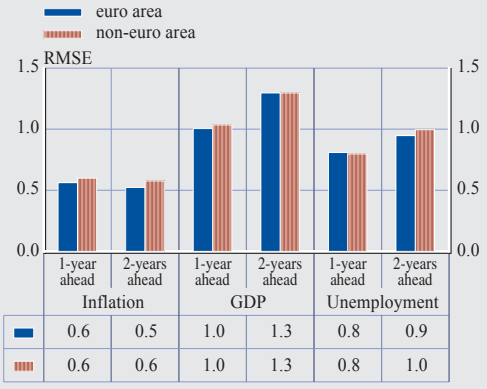
We now briefly consider whether the forecast performance of panel members differs according to their country of origin and, in particular, whether they are influenced by national economic developments when producing their euro area forecasts. For example, do panellists from countries with above-average inflation have systematically higher forecasts for inflation than panellists from countries with below-average inflation? Once again, the analysis should be interpreted with caution since, in some countries, only one or two respondents have responded a sufficient number of times to be scrutinised. Even in the case of Germany, which has the largest number of regular respondents, there are only between 8 and 11 forecasters depending on the variable and horizon in question.

Chart 7 indicates that for inflation and GDP growth there appears to be no relationship between the forecast error of respondents from a specific country and the average rate of growth of the forecast variable in the country.[29, 30] With regard to the unemployment rate, a small positive correlation exists, suggesting that countries with a lower unemployment rate are more likely to over-forecast the euro area unemployment rate; this is counter to the *a priori* expectation. A broadly similar picture holds true for the two-year ahead horizons. All in all, on the basis of this preliminary analysis, there appears to be no significant link between average national developments in the forecast variable and the forecast error for the euro area of respondents from each country. Chart 8 illustrates that there are no significant differences between average forecast errors (in terms of RMSE) of forecasters based inside the euro area and those based

29 Removing the outlier in terms of GDP growth (i.e. Ireland, where GDP growth averaged around 6% since 1999) yields a slightly significant negative relationship for the 4-quarter ahead horizon but not for the 8-quarter ahead horizon.

30 The same picture would hold true if one were to plot the relationship with the average forecast level of respondents from a specific country instead of the forecast error.

**Chart 8 Average forecast errors (RMSE) for forecasters based inside and outside the euro area**




| | Inflation 1-year ahead | Inflation 2-years ahead | GDP 1-year ahead | GDP 2-years ahead | Unemployment 1-year ahead | Unemployment 2-years ahead |
|---|---|---|---|---|---|---|
| euro area | 0.6 | 0.5 | 1.0 | 1.3 | 0.8 | 0.9 |
| non-euro area | 0.6 | 0.6 | 1.0 | 1.3 | 0.8 | 1.0 |

Source: ECB calculations.

outside the euro area (Denmark, Sweden and the United Kingdom). Although for four of the six permutations (three variables and two horizons), the average forecast error of forecasters inside the euro area is fractionally below that of the others – the exceptions being unemployment one year ahead and GDP two years ahead. This is, however, often only visible from the second decimal point.

### FINANCIAL AND NON-FINANCIAL INSTITUTIONS

Another issue that we investigate is whether forecasts from financial institutions are different from those from non-financial institutions (e.g. research institutes and employer or employee bodies). On the one hand, the commercial nature of financial institutions could mean that their credibility in forecasting is very important and that significant resources are therefore devoted to forecasting.[31] On the other hand, non-financial institutions may not face the same commercial pressures and disciplines that financial institutions face, but may also be more free to produce forecasts that deviate from the average. In any case, the evidence from Table 7 does not suggest any difference between these two broad types of forecasting institution. Across all the forecast variables and horizons, there are only minor differences between the different forecast statistics. Thus, in terms of the RMSE statistic, non-financial institutions did marginally better at forecasting inflation one year ahead, but slightly worse for unemployment two years ahead; for the other variables and horizons, the RMSE statistics were exactly equal. Conversely, financial institutions did marginally better (in terms of both mean error and mean absolute error) when forecasting GDP growth two years ahead. To sum up, the evidence to date does not suggest any significant difference in the average forecast performance of financial and non-financial institutions.

31 See Batchelor (2001) for a comparison of G7 economic forecasts made by private sector economists with those of intergovernmental agencies.

**Table 7 A comparison of the forecasting performance of financial and non-financial institutions**

| | | Inflation 1-year ahead | Inflation 2-year ahead | GDP 1-year ahead | GDP 2-year ahead | Unemployment 1-year ahead | Unemployment 2-year ahead |
|---|---|---|---|---|---|---|---|
| Mean error | Financial | 0.5 | 0.4 | -0.1 | -0.8 | -0.5 | -0.1 |
| | Non-financial | 0.5 | 0.4 | -0.1 | -0.9 | -0.4 | -0.1 |
| Mean abs. error | Financial | 0.5 | 0.5 | 0.8 | 1.1 | 0.6 | 0.7 |
| | Non-financial | 0.5 | 0.5 | 0.9 | 1.2 | 0.6 | 0.7 |
| RMSE | Financial | 0.6 | 0.5 | 1.0 | 1.3 | 0.8 | 0.9 |
| | Non-financial | 0.5 | 0.5 | 1.0 | 1.3 | 0.8 | 1.0 |
| Theil's U | Financial | 1.0 | 0.7 | 0.7 | 0.8 | 1.1 | 1.1 |
| | Non-financial | 0.9 | 0.7 | 0.8 | 0.8 | 1.0 | 1.1 |

Source: ECB calculations.

## 3 ASSESSING FORECAST UNCERTAINTY USING THE SPF

Although most economic forecasts are presented in terms of the point estimate of the variable being forecast, the point estimate itself conveys only a limited amount of information about the underlying forecast. As discussed in the introduction, the ECB SPF not only captures respondents' subjective assessments of the expected value for each of the three macroeconomic variables over alternative horizons, but also obtains specific quantitative information regarding the uncertainty surrounding these expectations. More specifically, SPF participants are also asked to assign a probability distribution to their forecasts. This distribution provides information about the probability, expressed as a percentage, of the future outcome being within a specific range. The probability distribution resulting from the aggregation of responses also helps to assess how, on average, survey participants gauge the risk of the actual outcome being above or below the most likely range. Such quantitative information on the uncertainty surrounding the macroeconomic outlook is increasingly highlighted as an important part of the overall macroeconomic information set that should be used by economic policy-makers and central banks (see, for example, the discussions in Tay and Wallis (2000), Wallis (2001) or Sims (2002)).[32, 33]

Measures of uncertainty, in general, may be useful for a number of reasons. First, they can provide additional insight beyond that provided by considering point estimates alone. For example, if a shock occurs, it might take some time for forecasters to assess the likely impact of this shock and to incorporate it into their baseline forecasts. However, they might first signal the possible impact of this shock in their estimation of uncertainty. It could also be the case that some risks, which are difficult to quantify, could give rise to changes in the assessment of uncertainty without necessarily changing the point forecasts. Second, uncertainty could be of interest in terms of considering credibility issues. In particular, although point estimates could be anchored by explicit policy goals, it is possible that changes in uncertainty could highlight an increase or decrease in confidence in the likelihood of the underlying goals being achieved.[34] Finally, uncertainty premia play an important role in many economic models of consumer and investor behaviour (see, for example, Lahiri et al. (1988), Giordani and Söderlind (2003) and D'Amico and Orphanides (2006)). Thus, movements in survey measures of uncertainty may help us to understand the evolution of financial market variables as well as households' and firms' decision-making. In addition, information on forecast uncertainty that is collected from private agents constitutes a useful assessment that may complement (though not replace) a central bank's own assessment of forecast uncertainty.

There are a number of ways forecast uncertainty may be measured in the context of the SPF. The best measure depends on the specific issue. For instance, one could be interested in the general level of uncertainty surrounding the point estimate, the probability that the outcome is above or below a certain threshold, or to what extent forecasters disagree with each other about the most likely outcome. The SPF allows us to address most of these aspects, as it provides us with information both on respondents' point estimates and on the probability distributions underlying their point estimates. Table 8 summarises each of the measures of forecast uncertainty considered in this section. We first consider "dispersion"

32 The importance of probability event and density forecasts can be justified within a decision-based framework (see Pesaran and Skouras (2002)). In particular, for users of forecasts, only in the restricted case of decision-making under quadratic loss functions and linear constraints are point forecasts sufficient for decision-making. Under more general non-quadratic loss functions and/or non-linear constraints, probability forecasts or even the entire predictive joint probability distribution of all relevant state variables are needed.

33 Zarnowitz and Lambros (1987) highlighted the relevance of the alternative measures of forecast uncertainty that can be constructed from survey data, while Giordani and Söderlind (2003) provide a more recent discussion.

34 This issue is addressed explicitly in Section 4.

**Table 8 Summary of forecast uncertainty measures**

| | Type | Use of point estimates or distribution | Calculation |
|---|---|---|---|
| 1 | Disagreement | Point estimates | Standard deviation of point estimates[1)] |
| 2 | "Individual uncertainty" | Distribution | Average standard deviation (variance) of individual distributions |
| 3 | "Aggregate uncertainty" | Distribution | Standard deviation (variance) of aggregate distribution |
| 4 | "Combined forecast uncertainty" | Distribution and point estimates | "Combined forecast" standard deviation (variance)[2)] |
| 5 | "Proxy skew" | Distribution and point estimates | Mean of aggregate distribution minus average of point estimates |
| 6 | "Individual kurtosis" | Distribution | Average kurtosis of individual distributions |
| 7 | "Aggregate kurtosis" | Distribution | Kurtosis of aggregate distribution |

1) Use either unadjusted or quasi-standard deviation.
2) Proxied by difference between individual uncertainty (i.e. average variance of the individual distributions) and disagreement (i.e. variance of point estimates).

(both in terms of point estimates and probability distributions) in the measures of uncertainty (numbers 1-4 in Table 8) before considering other characteristics of the forecast probability distribution that may also be informative.

## 3.1 ALTERNATIVE DISPERSION-BASED UNCERTAINTY INDICATORS FROM THE SPF

### DISAGREEMENT AS A MEASURE OF UNCERTAINTY

One of the most frequently used approaches to measuring uncertainty in surveys has been to look at disagreement (or consensus) among survey respondents about their point forecasts (see Zarnowitz and Lambros (1987)).[35] Using disagreement as a measure of uncertainty makes some sense, to the extent that different forecasters have a different assessment of the outlook, which, in turn, reflects the overall uncertainty surrounding the outlook. Hence, a high level of disagreement may be indicative of a high level of uncertainty. However, a number of caveats should be borne in mind. First and foremost, disagreement is not a pure measure of uncertainty but only a proxy. As a result, the level of disagreement may not correlate perfectly with actual forecast uncertainty. For example, even though the overall level of uncertainty may have increased, it is quite possible that forecasters may at the same time increasingly agree with each other about the most likely outcome. It could also be that forecasters do not change their view on the most likely outcome, but feel more or less uncertain about it. In such cases, indicators of disagreement cannot capture the change in uncertainty and so an alternative measure is required. Second, the scope for disagreement may be related to the number of respondents. Third, if there are differences in the quality of the forecasters or in the frequency or timing with which forecasters update their forecasts, an outlying forecast may not be indicative of disagreement or uncertainty, but may simply indicate that a specific forecaster has not taken relevant information on board.[36]

### DISPERSION INDICATORS DERIVED FROM THE FORECAST DISTRIBUTION

If there is information on the probability distribution underlying the point forecasts, a more natural measure of uncertainty is the spread of the forecast distribution, as measured

35 Disagreement is usually measured by the standard deviation of the point forecasts, which captures the dispersion of point estimates across the panel of participants. Sometimes, in order to reduce the possible impact of outliers, the quasi-standard deviation is used (see Giordani and Söderlind (2003)). The quasi-standard deviation proxies the standard deviation by taking half the difference between the 84th and 16th percentiles (i.e. 68% of the distribution).

36 Although the SPF is made up of professional forecasters, it is possible that forecasters do not update the forecasts at the same time or with the same frequency.

by its standard deviation (or variance).[37] The standard deviation of a distribution is essentially a summary statistic of the dispersion of possible events around the mean. Thus, the more it is considered that possible future events are dispersed from the mean, the higher the standard deviation and the associated future uncertainty. However, different interpretations may be drawn depending on whether one considers the average standard deviation of each individual probability distribution or the aggregate distribution.

The average standard deviation of the individual distributions captures uncertainty as reported by individual respondents, but does not in any way capture differences across individuals. Thus, it is possible that individuals report the same standard deviation from one round to the next, but disagree more among themselves about the actual point estimate. In this case, simply taking the average of the individual standard deviations does not capture the changes in the dispersion of the point estimates, even though such increased disagreement may partly reflect a change in overall uncertainty.

A second approach is to take the standard deviation of the aggregate distribution.[38] It is possible to show that the variance of the aggregate distribution is equal to the average variance of the individual distributions (i.e. individuals' uncertainty) plus the variance of the point estimates (i.e. disagreement). Thus, this measure takes into account both individual uncertainty and disagreement.[39]

### COMBINED FORECAST UNCERTAINTY

A third possible measure of forecast uncertainty that can be constructed using the SPF data is what Giordani and Söderlind (2003) call the "combined forecast" uncertainty. They argue that if the sample is large enough, cross-sectional idiosyncratic movements in the forecast errors are averaged out and one is left with only common movements. They also show that, under certain circumstances, the combined forecast uncertainty may be written as the average variance of the individual distributions (i.e. individuals' uncertainty) minus the variance of the point estimates (i.e. disagreement). Thus, when forecasters disagree, combined forecast uncertainty will be smaller than the uncertainty measured by the average spread of individual forecast distributions. Such a measure may be particularly useful in the context of the SPF in order to provide an aggregate uncertainty indicator which takes account of the possible reduction in uncertainty due to the pooling of individual information sets. However, one drawback of this measure is that it is possible for the proxy of combined forecast variance to be negative if the variance of the point estimates is larger than the average variance of the individual distribution. Whilst practically this is not an insurmountable problem, theoretically the variance cannot be negative. Box 3 provides an example which highlights the advantages and disadvantages of the combined forecast uncertainty measure and the measures based on the aggregate and individual distributions.

37 There are two ways of presenting the dispersion of a distribution, either in terms of its standard deviation or its variance (the standard deviation is simply the square root of the variance). The variance is useful as the different measures of dispersion presented below are linked to each other in terms of variance. However, the standard deviation (which is simply the square root of the variance) is usually reported, as it is in the same "dimension" (e.g. percentage points, millions) as the underlying variable. Furthermore, the standard deviation is often used when calculating confidence intervals for distributions.

38 The aggregate distribution is constructed simply by summing up the individual probabilities reported in the SPF and dividing by the number of respondents. Note that this "aggregate" distribution which combines the individual density forecasts (or probability distributions) is a specific form of what is known as a finite mixture distribution (see Wallis (2004)).

39 For a further discussion on the relationship between disagreement, individual uncertainty and aggregate uncertainty, see Giordani and Söderlind (2003) and Linden (2003).

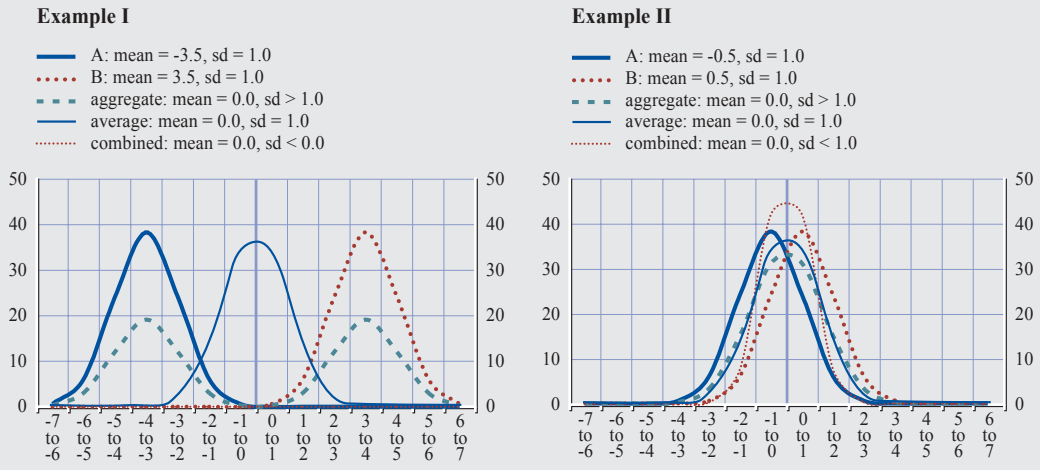
Box 3

## POTENTIAL LIMITATIONS OF THE ALTERNATIVE FORECAST UNCERTAINTY MEASURES

The chart below shows a simple illustration, which may serve to highlight the intuition behind and possible drawbacks of the alternative uncertainty measures. In this illustration, we have just two individuals (A and B), who report distributions with differing means (-3.5 and 3.5 in Example I and -0.5 and 0.5 in Example II) but the same standard deviations (unity). If one was to consider the simple average of the mean and standard deviation of the two distributions, the result would be a mean of zero and a standard deviation of unity in both examples. The aggregate distribution, however, whilst also having a mean of zero, would have a standard deviation larger than unity in both examples. This is because the disagreement between the two individuals is also taken into account. In this example, the combined forecast would also have a mean of zero, but in Example I the distribution is not defined, as the disagreement between the two individuals is larger than their individual assessment of uncertainty. Whilst the example presented may be an extreme case, it does occur in practice that the variance of the point estimates (or disagreement) is larger than the average individual assessment of uncertainty (the average variance of the distributions) and hence the combined forecast distribution is not defined. This is particularly the case for long-term unemployment expectations. In practice, this may indicate some problems with the assumptions required to reach the simplified measure of combined forecast uncertainty mentioned above. In Example II, the combined forecast uncertainty is defined and is lower than that of either the average or aggregate distributions.

### Alternative measures of forecast uncertainty

(percentages)




Note: sd denotes standard deviation. Average refers to the average of the individual distributions, aggregate to the aggregate distribution and combined to the combined forecast".

To sum up, different indicators of uncertainty can therefore be constructed by considering the information contained in the individual distributions from the SPF in different ways. Each measure has some justification and may provide a different insight into the overall level of uncertainty surrounding the SPF expectations. It is therefore useful to compare and contrast the different measures in practice.

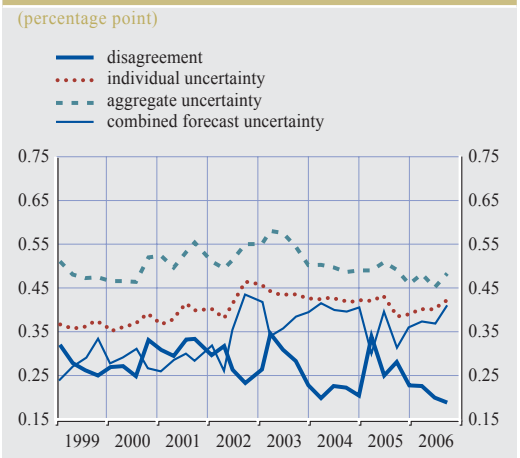
### 3.2 COMPARISON OF ALTERNATIVE UNCERTAINTY INDICATORS FROM THE SPF

It should be noted that calculating the standard deviation of the SPF distributions is somewhat problematic in practice. The reason for this is that respondents are asked to attach probabilities to specific ranges. Thus, we only have information on the probability attached to the outcome being in a specific range, and not on how the probability is distributed within that range. If one assumes that all the probability is located in the centre of the range, this may bias upward the estimated standard deviation, as it is more likely that more of the probability within a range is located closer to the centre of the distribution than further away.[40]

One solution to the problem is to fit a Normal distribution to the individual distributions provided by SPF respondents (see, for example, D'Amico and Orphanides (2006));[41] this will yield an estimated mean and a less biased estimate of the standard deviation. However, in some cases, the assumption of normality appears too extreme and also gives rise to other problems (such as eliminating skew and kurtosis). Hence, in what follows, we use an approach which assumes that all the probability for a given range relates to the mid-point of that range.[42]

Chart 9 plots the different indicators of uncertainty mentioned above for the case of one-year ahead inflation forecasts. The chart highlights the possible differing messages depending on which aspect of uncertainty is being focused on. Disagreement about point forecasts one year ahead, as measured by the standard deviation of the point estimates, fluctuated around 0.3 percentage point for most of the SPF rounds until 2003. Since then, however, there has been a decline in the level of disagreement to approximately 0.2 percentage point. Considering that the standard deviation of actual annual inflation rates and average forecast errors since 1999 have been around 0.5 percentage point, this level of disagreement is quite low, which, together with a gradually declining profile over the latter part of the sample period, suggests that it may not be a good proxy for overall inflation uncertainty; although the standard deviation of a series in itself does not indicate the uncertainty surrounding forecasts of that series. As regards the standard deviation of the aggregate distribution, this has tended to be at or slightly above 0.5 percentage point, although it was slightly lower between 1999 and 2001 and has been since 2004. The fact that the standard deviation of the aggregate distribution corresponds broadly to the level of historical

**Chart 9 Short-term inflation forecast uncertainty**




Source: ECB calculations.
Note: Disagreement is calculated as the standard deviation of the point estimates. Individual uncertainty is calculated as the average standard deviation of the individual probability distributions. Aggregate uncertainty is calculated as the standard deviation of the aggregate probability distribution and is positively related to both individual uncertainty and to disagreement. The combined forecast uncertainty is positively related to individual uncertainty and negatively related to disagreement.

40 One approach to correcting for this likely bias is to apply "Sheppard's correction" (Stuart and Ord (1994)), which adjusts the calculated variance by subtracting one-twelfth of the squared bin width. However, this correction is not without its limitations either, particularly in the context of the SPF; see, for example, Lahiri and Liu (2006).

41 When normal is written with a capital letter (i.e. Normal), it is to denote that it is being used in the statistical sense of the word (e.g. that the distribution is Normally distributed).

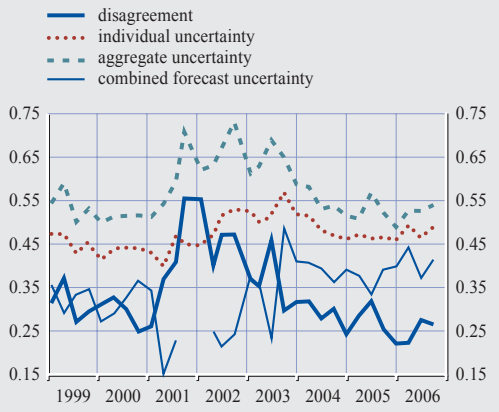
42 As a cross-check, when Normal distributions were fitted to the aggregate and individual SPF distributions, the results and conclusions reported here still held; in some cases more strongly. For example, as the standard deviations of the Normal distributions are generally smaller than when it is assumed that all the probability relates to the mid-point of each range, this implies that the finding that SPF respondents generally underestimated uncertainty is robust.

inflation volatility suggests that it may be a more reliable measure of the uncertainty surrounding the inflation outlook. The level of average individual uncertainty has been significantly lower, also compared with the historical standard deviation of actual inflation, suggesting that individual panel members may not fully capture the overall level of inflation uncertainty.[43] It is, however, interesting to note that the profile of this indicator is somewhat different as it has tended to edge upwards over the period in question, suggesting that respondents have gradually adapted their perceptions to be more in line with the actual outcome. The level of uncertainty, as measured by the combined forecast, has also tended to edge upwards, driven by a combination of the lower level of disagreement and higher average individual uncertainty. Notwithstanding this gradual upward movement, the combined forecast uncertainty remains below the level that would be implied by the historical standard deviation of inflation over the period.

The level of disagreement among SPF forecasters concerning the short-term outlook for GDP has been higher on average and more volatile than that for inflation. In particular, whilst it averaged about 0.3 percentage point between 1999 and 2001, after the terrorist attacks in the United States on 11 September 2001, it rose sharply, reaching over 0.5 percentage point at the end of 2001 (see Chart 10). Given that the actual standard deviations of the annual GDP growth rate and average forecast error have been greater than one percentage point since 1999, this level of disagreement is quite low. The profile of the standard deviation of the aggregate distribution has followed a similar pattern, but also at a relatively low level compared with the actual volatility of GDP growth. The profile of the combined forecast is not defined around end-2001 and the beginning of 2002 as disagreement between forecasters was larger than the average individual assessment of uncertainty. At around 0.45 percentage point, average individual uncertainty has been the most stable, although it did edge up between mid-2001 and late-2003, before easing back somewhat more recently.

**Chart 10 Short-term real GDP growth forecast uncertainty**

(percentage point)



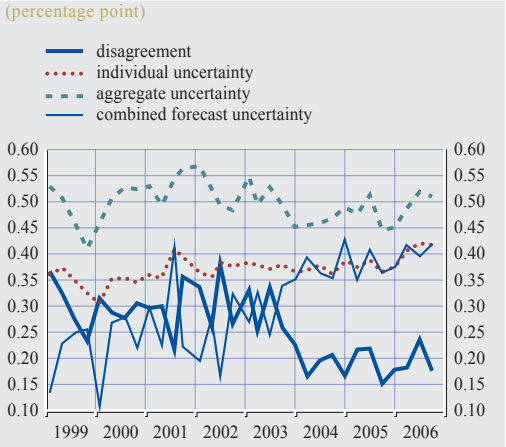

Source: ECB calculations.
Note: Disagreement is calculated as the standard deviation of the point estimates. Individual uncertainty is calculated as the average standard deviation of the individual probability distributions. Aggregate uncertainty is calculated as the standard deviation of the aggregate probability distribution and is positively related to both individual uncertainty and to disagreement. The combined forecast uncertainty is positively related to individual uncertainty and negatively related to disagreement.

The profile of the level of disagreement about the short-term unemployment rate forecast was broadly similar to that of the short-term inflation forecast, i.e. fluctuating around 0.3 percentage point for most of the period since 1999, but has declined since 2004 (see Chart 11). The standard deviation of the unemployment rate has been 0.5 percentage point over the same period. As regards the profile of the standard deviation of the aggregate distribution, there has been no clear pattern over time, with the exception that it has also tended to fluctuate around 0.5 percentage point on average.

To sum up, the evolution of the different measures of uncertainty has differed both according to the measure used and to the variable being considered. Based on the above analysis, our preferred measure of uncertainty is the standard deviation of the aggregate probability distribution, as it combines information both on the amount of

43 This is particularly the case when one allows for the upward bias in the calculated measure of the standard deviation of the distribution from assuming the weight is located in the centre of the ranges.

**Chart 11 Short-term real unemployment rate forecast uncertainty**




Source: ECB calculations.
Note: Disagreement is calculated as the standard deviation of the point estimates. Individual uncertainty is calculated as the average standard deviation of the individual probability distributions. Aggregate uncertainty is calculated as the standard deviation of the aggregate probability distribution and is positively related to both individual uncertainty and to disagreement. The combined forecast uncertainty is positively related to individual uncertainty and negatively related to disagreement.

*disagreement between forecasters* and on their *individual assessment of uncertainty*. In the case of inflation, it is also the measure most consistent with the actual historical volatility of the forecast series (inflation) over the period. At the same time, caution is warranted when using the above indicators to assess macroeconomic risks. In particular, at the individual level, SPF respondents do not seem to fully capture the overall level of macroeconomic uncertainty. Moreover, even at the aggregate level, a more sophisticated evaluation of the SPF density forecasts (see Annex 3) suggests some evidence that respondents failed to accurately assess the risk of higher inflation outcomes over this period. In addition, SPF respondents may have also failed to accurately assess the risks of lower growth and higher unemployment associated with the economic downturn.

### 3.3 OTHER CHARACTERISTICS OF THE FORECAST PROBABILITY DISTRIBUTION

If the probability distribution is Normally distributed, the mean and the standard deviation are sufficient to capture all the information contained in that distribution. However, as we shall see below, the SPF probability distributions (both at the individual level and aggregate level) frequently deviate from normality.[44] As a result, it is important to consider other summary statistics in order to fully describe the information on macroeconomic uncertainty contained in the SPF probability distributions. In particular, higher "moment statistics", such as the skew and kurtosis, may also be relevant.

The skew is a measure of asymmetry in the distribution, i.e. whether the distribution looks the same to the right and the left of the mean.[45] Knowledge about the skew of a distribution may be particularly relevant where asymmetric risks are considered particularly important. For example, if an outcome of higher inflation is considered more damaging than an outcome of lower inflation, financial market participants attempt to hedge against inflation risk.

An alternative measure of skew, which might be more robust to the calculation problem (see footnote 45), is to consider the relationship between the mean of the distribution and the point forecast as reported by respondents. To the extent that the point forecast can be interpreted as a modal or most likely value, the difference between the mean and the mode can

44 Lahiri et al. (1988) show that the probability distributions underlying forecasts may not be Normal for the United States.

45 In practice, calculating the skew of the SPF distributions is somewhat problematic. The main reason for this is that respondents are asked to attach probabilities to specific ranges. Thus, we only have information on the probability attached to the outcome being in a specific range, and not on how the probability is distributed within that range. If it is assumed that all the probability is located at the centre of the range, this distorts the estimated skew, as it is more likely that more of the probability associated within a range is located closer to the centre of the distribution than further away. The possible solution cited above for estimating the standard deviation (i.e. by fitting a Normal distribution to the individual distributions provided by SPF respondents and maximising the fit over different estimates of the mean and standard deviation) is not suitable in this case, as fitting a Normal distribution suffers from the drawback that it, by definition, removes any skew and excess kurtosis at the level of the individual distributions. Thus, higher moment information is lost. It is also possible to fit alternative classes of distributions to the observed distribution. However, this adds another element of uncertainty to the interpretation of the distribution.

be used as an indicator of skew.[46] In particular, if the distribution is symmetric, the mean and modal forecast should be equal. However, if the distribution is skewed in either direction the two will differ. Thus, if there is upward skew (i.e. the risks are skewed to the upside), the mean will be larger than the mode and vice versa.

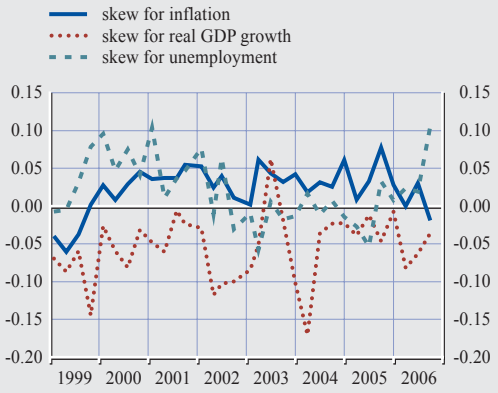
Another specific way of considering uncertainty is to examine the probability associated with an outcome being above or below a specific threshold or value. This approach could be relevant when there is a policy target or officially declared range for the long-term value of a variable. For example, the ECB Governing Council announced that, over the medium term, inflation should be below, but close to, 2%. It is possible that market participants believe that this is the most likely outcome but may change their view on the risks associated with a higher outcome. Thus, the probability associated with an outcome above 2% might be a useful indicator of the upside risk that forecasters associate with the ECB failing to meet its policy aim. Such a risk indicator is discussed in more detail in Section 4 on long-term expectations.

The kurtosis of a distribution is often described as the "peakiness" of the distribution. It is a measure of whether the data are peaked or flat relative to a Normal distribution (i.e. excess kurtosis). Although the kurtosis of a distribution may also contain useful information about uncertainty, in particular, the likelihood of extreme events relative to a Normal distribution, in practice calculating the kurtosis of the SPF distributions is affected by similar problems to those outlined for calculating the skew.[47] Nonetheless, since the weights in the tails of the aggregate distribution will be larger, the more disagreement there is among forecasters, one might expect, a priori, the degree of kurtosis of the aggregate distribution to be larger than that of the individual distributions.

Considering the skew measures in practice, it appears that the balance of risks varies considerably across variables and over time in the SPF. As regards the balance of risks to the short-term inflation outlook, it appears that, with the exception of the earlier and latest rounds, SPF forecasters believe that they have been on the upside (see Chart 12). Interestingly, whilst respondents correctly perceived the balance of risks to be on the upside, they may have underestimated the magnitude of these risks. Indeed, since 2000, the mean of the aggregate distribution has been approximately 0.03 percentage point above the average of the point estimates.[48] However, the actual outcome has been even higher, with the mean error being 0.5 percentage point. Similarly, although actual HICP annual inflation since 2000 has averaged above 2%, until the round for the second quarter of 2006, the aggregate probability associated with an outcome above 2% was never more than 50%, even though it did increase substantially over the course of 1999 and 2000.

**Chart 12 Indicators of skew surrounding short-term one-year ahead forecasts**

(proxied by the mean of the aggregate distributions minus the average of point forecasts)



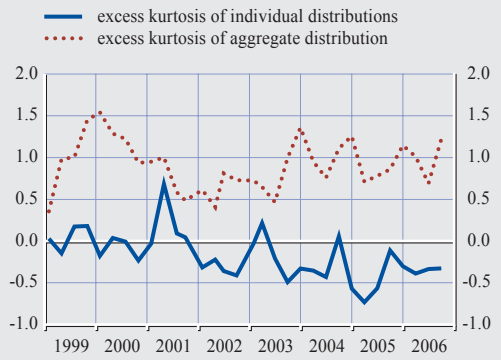

Source: ECB calculations.

46 It is not always clear whether respondents are providing a mean (most likely on average), mode (most likely outcome) or median (outcome that has 50% of the distribution above and below) forecast.

47 Possible proxy measures include considering the kurtosis of the distribution of point estimates or considering the weight of the distribution in the tail ranges.

48 We calculate the mean of the distribution by taking the mid-point of each range (e.g. for the range 1.5-1.9% the mid-point is 1.7%) and multiplying it by the probability reported by respondents of the outcome being in that range.

**Chart 13 Kurtosis of short-term (one-year ahead) inflation forecasts**




Source: ECB calculations.

In contrast to the short-term inflation forecasts, the skew associated with the short-term GDP forecasts has generally been to the downside. As regards the skew of the unemployment distribution, the mean at the beginning of the sample period was generally above the average of the point estimates. This would suggest that the SPF forecasters perceived upward risks, although since 2002 the balance of risks has been broadly balanced, with the exception of the latest round in the fourth quarter of 2006. As this lower degree of upward skew has occurred after a period of gradual upward revisions in unemployment rate expectations, it suggests that some of the higher upside risks in the early part of the sample were gradually incorporated into respondents baseline expectations.

Chart 13 plots the excess kurtosis measures for the one-year ahead inflation forecast. Two measures are presented, one based on the average excess kurtosis (compared with a Normal distribution) of the individual distributions and one based on the aggregate distribution. A general result for each variable and time horizon is that the average of the former is usually smaller than the latter. For example, Chart 13 shows the individual and aggregate measures of kurtosis for the short-term inflation forecasts. This result is not surprising, as the aggregate distribution is affected by the differing means (i.e. disagreement) of the individual distributions, tending to add more probability mass to the tails of the aggregate distribution. Thus, on average, and relative to a Normal distribution, the individual distributions attach less weight to extreme outcomes. Over the first half of the sample period, the average individual distribution attached a similar weight to extreme events as would be predicted by a Normal distribution, i.e. the excess kurtosis hovered around zero. By contrast, the aggregate distribution attaches more weight to extreme events relative to a Normal distribution.

## 3.4 OVERALL ASSESSMENT OF UNCERTAINTY MEASURES

Table 9 presents a summary of the average values of some of the alternative measures of uncertainty for the short-term horizons that can be calculated from the SPF data. Although these measures are not exhaustive, they should cover many of the different aspects of uncertainty that one might wish to address. The first measure (disagreement) does not capture uncertainty in terms of probability, but may nonetheless be an informative proxy and is easily understood. The different measures of the standard deviation/variance of the distribution are closer to the classical concept of uncertainty, although the different measures can give rise to different assessments of the level of uncertainty. The skew and kurtosis measures give some insight into uncertainty, either in a directional sense (i.e. a preponderance of upside or downside risks) or relative to the mean (i.e. are extreme events relatively likely or not).

Although it is clear that uncertainty is a multi-faceted concept that does not lend itself to easy summary by one statistic, there is a danger of confusion with such a range of statistics. Therefore, although we compute and report each of the different measures, we would suggest that the standard deviation of the aggregate distribution should be taken as a benchmark measure from the SPF. In this regard, the fact that the aggregate uncertainty

**Table 9 Summary of the average values of uncertainty measures for short-term horizons since SPF round for the first quarter of 1999**

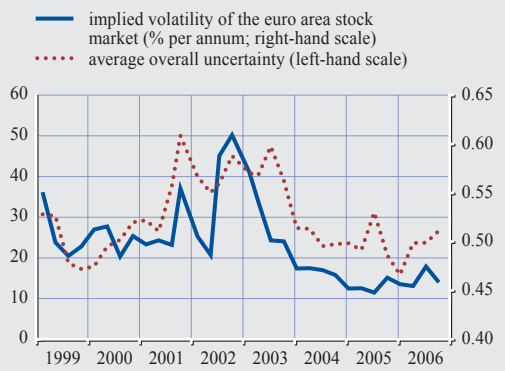
| | Type | Inflation [1] | GDP growth [1] | Unemployment [1] |
|---|---|---|---|---|
| 1 | "Disagreement" | 0.27 | 0.33 | 0.26 |
| 2 | "Individual uncertainty" | 0.40 | 0.47 | 0.37 |
| 3 | "Aggregate uncertainty" | 0.50 | 0.57 | 0.50 |
| 4 | "Combined forecast uncertainty" | 0.34 | 0.34 | 0.30 |
| 5 | "Proxy skew" | 0.03 | -0.06 | 0.02 |
| 6 | "Individual kurtosis" | -0.17 | -0.32 | -0.12 |
| 7 | "Aggregate kurtosis" | 0.91 | 1.25 | 1.89 |

Source: ECB calculations.
1) Average value since the first quarter of 1999 SPF round for rolling one-year ahead horizon.

measures are lower than the forecast RMSE statistics for each variable supports the assessment of the analysis in Annex 3 that respondents failed to accurately assess the risks over the period under consideration.

In this regard, it is interesting to consider whether this measure of uncertainty tallies with our qualitative assessment of macroeconomic uncertainty. Since 1999, it is thought that geo-political events have impacted significantly on the overall uncertainty surrounding economic forecasts. In particular, the terrorist attacks in the United States on 11 September 2001 and subsequent developments are thought to have considerably heightened the uncertainty and volatility in key macroeconomic variables, such as oil prices and stock markets. It may therefore be interesting to construct an aggregate indicator of macroeconomic uncertainty based on the three variables included in the SPF. For example, Chart 14 shows a simple arithmetic average of the standard deviations of aggregate distributions for the short-term forecast for each of the key variables in the SPF (inflation, growth and unemployment). Although this is a crude summary statistic, it appears to show that "average" uncertainty was indeed higher between the second half of 2001 and 2003. Since 2004, however, "average" uncertainty appears to have returned to more normal levels. The indicator also correlates quite well with an indicator of stock market volatility implied from option prices. In this regard, a promising area for future research will be to understand more fully the factors driving the uncertainty inherent in economic forecasts and expectations.

**Chart 14 Average of uncertainty surrounding short-term forecasts of inflation, growth and unemployment, as well as implied stock market volatility**




Sources: Bloomberg and ECB calculations.

## 4 SPF LONG-TERM EXPECTATIONS

SPF panellists are also invited to provide an assessment of their longer-term outlook for the euro area economy (i.e. five-year ahead forecasts for inflation, real GDP growth and unemployment in the euro area). By nature, these longer-term forecasts are expected to be free of the influence of cyclical developments and various shocks that are generally not foreseeable at this horizon. Longer-term forecasts might thus be expected to be less volatile than the shorter-term forecasts and to provide a picture that yields some insight to participants' views of the main trends that are thought to shape the euro area economy in the longer term. For example, longer-term inflation

forecasts may help to assess the ECB's credibility in maintaining the objective of price stability; longer-term real GDP growth forecasts may provide an insight into the expected trend potential growth for the economy and the longer-term unemployment expectations may provide an assessment of market participants' perception of the possible effects of structural reforms and changes in the labour market.

There are fewer forecasts available for longer-term horizons than for shorter-term horizons. The question about longer-term expectations was only asked once in 1999 and once in 2000 (in the January surveys). From the beginning of 2001, it was then included in every survey round. The maximum number of replies is therefore limited to 26 survey rounds compared with 32 for the shorter horizons. Furthermore, longer-term expectations are generally provided by a smaller number of participants. As we have seen, the SPF collects the views of 59 forecasters on average in each round. Of these, an average of 45 provided long-term point expectations and 39 added probability distributions to these long-term forecasts. Inflation and GDP growth have slightly higher response rates than unemployment for which only 41 panellists on average provide a long-term forecast. Assessing and interpreting SPF replies about longer-term expectations might thus warrant greater caution given the smaller number of observations. Nevertheless, around three-quarters of those responding to the survey do provide long-term expectations and this share has been quite stable since the beginning of the survey, providing useful information on expectations for the long-term outlook for the euro area economy.[49]

### 4.1 LONG-TERM INFLATION EXPECTATIONS: AN INDICATOR OF THE ECB'S CREDIBILITY

The ECB's monetary strategy aims to ensure long-term price stability, defined, in quantitative terms, as a "year-on-year increase in the HICP for the euro area of below, but close to, 2%". Well-anchored long-term inflation expectations are thus a desirable feature in the monetary policy framework, since they reflect the extent to which economic agents believe the ECB will achieve this inflation objective. Long-term inflation expectations may be derived from financial markets either from the term structure of interest rates or from the difference in yields between conventional Treasury bonds and bonds specifically linked to the inflation rate.[50] None of these measures is ideal since they may be affected by technical market factors and various premia, in particular an inflation risk premium. Direct long-term expectations obtained from surveys, such as the SPF, may therefore provide a valuable complementary piece of information.

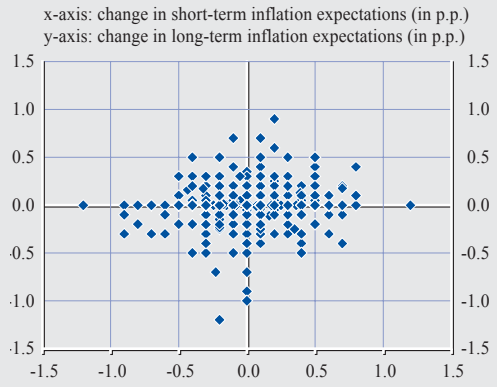
A central bank's ability to anchor inflation expectations can be assessed by the extent to which long-term inflation expectations respond to changes in very short-term inflation expectations (see, for example, Castelnuovo et al. (2003), Bordes and Clerc (2004)). In other words, if a shock, such as a surge in oil prices, pushes up the short-term inflation outlook but has no visible effect on long-term inflation expectations, it means that economic agents believe that the central bank will ensure that such developments will not spill over into more medium-term price pressures. Using SPF replies, one way to assess the credibility of monetary policy is to check whether there is no (or only very little) correlation between changes in short-term and longer-term inflation expectations. Chart 15 plots the individual revisions in short term (i.e. one-year ahead) and in the long term (i.e. five-year ahead) inflation expectations for the period from the first quarter of 1999 to the fourth quarter of 2006. It suggests that there is indeed no more than a weak relationship between movements in short-term and long-term inflation forecasts. The correlation between the two is statistically insignificant at 0.08, suggesting that the ECB has been fairly successful in anchoring long-

49 The number of respondents to the questions on longer-term expectations also compares favourably with Consensus Economics and the Euro Zone Barometer, both of which have approximately ten respondents replying to their questions on longer-term expectations.

50 For more details on the derivation of the "break-even inflation rate" derived from financial markets, see ECB (2002).

**Chart 15 Relationship between revisions to short-term and long-term inflation forecasts**

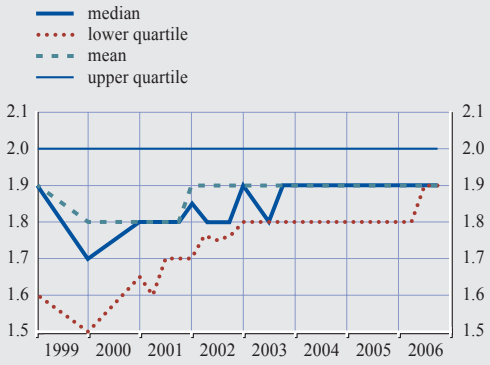
(Q1 1999-Q4 2006; individual observations)




Source: ECB calculations.

**Chart 16 SPF long-term inflation forecasts**

(percentages)




Sources: ECB and ECB calculations

term expectations, particularly compared with other central banks (see Bordes and Clerc, 2004).

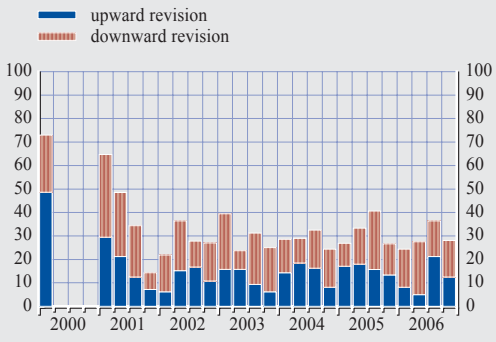
Castelnuovo et al. (2003) add another condition for long-term expectations to be considered well anchored. Long-term forecasts must exhibit a low level of volatility around a level compatible with the target range defined by the quantitative announcement of the inflation objective. Chart 16 reports the developments of long-term inflation forecasts made by SPF participants since the beginning of the survey. It shows that average long-term expected inflation has remained quite stable since the beginning of the survey. On average, it stood at 1.88% with a standard deviation of ±0.04 percentage point. The average long-term inflation expectation was 1.9% at the start of Stage III of EMU in 1999. It declined to 1.8% in 2000 and then shifted upwards to stand at 1.9% again at the end of 2002. Since then, it has remained broadly stable at below, but close to, 2%, confirming the stability of SPF long-term inflation expectations.

Chart 16 also plots the lower and upper quartiles, which correspond to the 25th and the 75th percentile responses. The upper quartile has remained unchanged at 2.0% since the beginning of the survey. In other words, at least 25% of forecasters have always expected long-term inflation to be equal to or above 2%. The lower quartile, on the other hand, increased steadily until the first quarter of 2003 and remained flat at 1.8% until the third quarter of 2006 when it increased further to 1.9%, indicating that at least 50% of SPF forecasters believed long-term inflation would stand between 1.9% and 2.0%. This range has narrowed since the beginning of the survey and has remained relatively stable since the review and clarification of the ECB's monetary policy strategy (in the second quarter of 2003), when the ECB announced that price stability was defined as inflation not only being "below", but also "close to", 2%.

Shifts in the median, also reported in Chart 16, reveal that this growing consensus does not imply that SPF forecasters take the long-term ECB inflation objective for granted without questioning it. A significant number of individual responses are indeed revised from one survey round to the next. When the SPF survey first started, a particularly large number of respondents (around two-thirds in the first quarters of 2000 and 2001) changed their forecast for long-term inflation. This may reflect a greater understanding of the ECB's

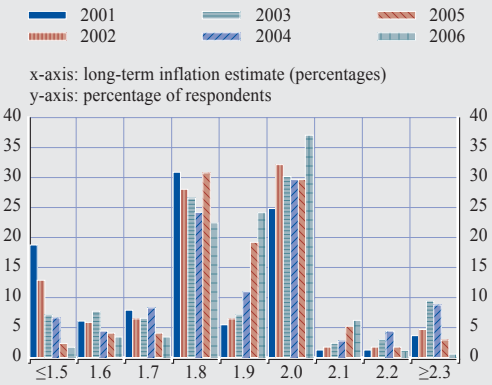
**Chart 17 Percentage of SPF participants revising their long-term inflation expectations from one round to the next**

(as a percentage of participants having provided a reply in two consecutive rounds)




Source: ECB calculations.

**Chart 18 Distribution of long-term inflation forecasts among SPF respondents**




Source: ECB calculations.

medium-term monetary policy objective. Since 2001 this number has fallen, but remains significant at around 30% for each survey round since 2002 (see Chart 17). Overall, since the beginning of the survey, nearly all SPF participants who have provided long-term inflation expectations have adjusted their expectations at least once. Among those, about two-thirds have tended to increase their expectations (by an average 0.3 percentage point) and the remaining third has tended to reduce their expectations by about the same amount.[51]

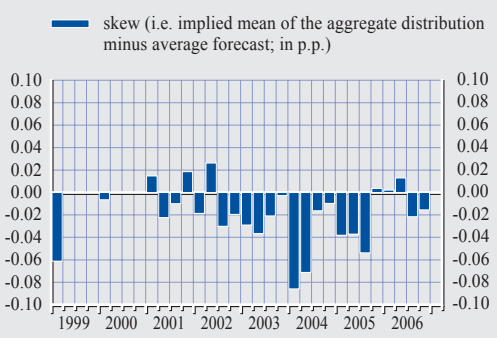
Consequently, the distribution of long-term inflation estimates shifted from survey to survey, reflecting the changes of SPF respondents' forecasts (see Chart 18). In particular, the number of SPF respondents providing long-term inflation forecasts below 1.8% decreased substantially between 2001 and mid-2005: in 2001 more than 30% of respondents provided a long-term inflation forecast below 1.8%, whereas in 2006 only 8% did so. This illustrates both the changes in opinion regarding the long-term inflation outlook and the gradual narrowing in the range of replies provided by SPF respondents. It is also interesting to see that SPF long-term inflation forecasts have mainly tended to be either 1.8% or 2.0% and that, until 2005, relatively few respondents projected long-term inflation at 1.9%. While it is impossible to know the reasons behind these revisions of longer-term inflation expectations, the gradual decline in SPF respondents' disagreement might reflect the impact of the ECB's clarification that price stability is considered consistent with HICP inflation at below, but close to, 2% over the medium term.

The average probability distribution has not changed dramatically since the beginning of the SPF survey. In particular, there has been a consistently low perceived risk of deflation. In addition, there has always been less than a 10% probability assigned to the risk of inflation being below 1%. In general, SPF respondents believe that the probability that long-term inflation will eventually stand between 1.5% and 1.9% is the highest (38% on average), in line with their average point estimate. Although this probability tended to decrease slowly between 2001 and 2004, it has rebounded somewhat since the beginning of 2005.

A measure of the skew of the probability distribution, derived as the difference between the implied mean of the aggregate distribution

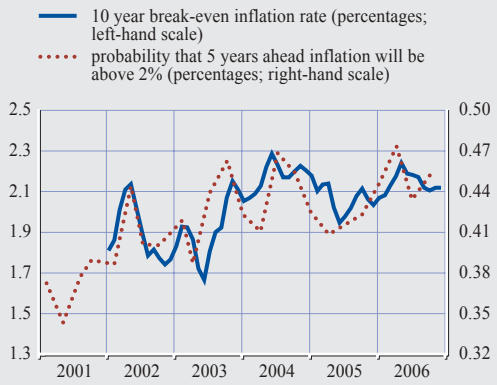
51 These estimates correspond to the difference between the average long-term expectation in the first year of participation (generally 1999) and that of the last year of participation (generally 2006). It does not take into account possible upward or downward revisions in between. Annual individual averages are rounded to the first decimal point.

**Chart 19 The balance of risks surrounding long-term inflation expectations**




Source: ECB calculations.
Note: there were no questions related to longer-term inflation forecasts in the 1999 Q2, 1999 Q3, 1999 Q4, 2000 Q2, 2000 Q3 and 2000 Q4 survey rounds.

**Chart 20 Probability that five-year ahead HICP inflation may be equal to or above 2% and the ten-year break-even inflation rate**



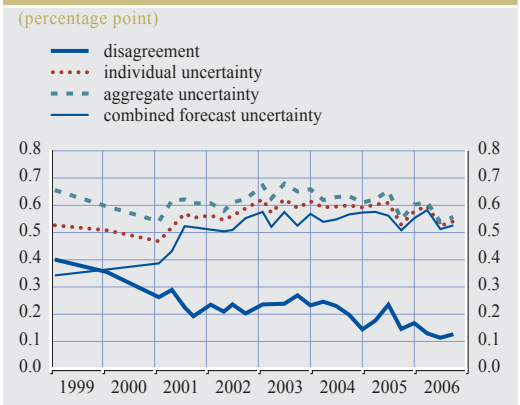

Sources: Reuters and ECB calculations.
Note: Break-even inflation rates correspond to monthly averages; SPF five-year ahead inflation forecast corresponds to SPF panellists' expectations around the middle of the month surveyed.

and the average long-term inflation forecast, provides a clearer picture of the balance of risks attached to long-term inflation. Chart 19 shows that the overall balance of risks attached to long-term inflation has been slightly on the downside on average. Note, however, that we calculate the mean of the distribution by taking the mid-point of each range (e.g. for the 1.5%-1.9% range the mid-point is 1.7%) and multiplying it by the probability of the outcome being in that range, as reported by the respondents. It could also be argued that the mid-point in the example above is 1.75% (i.e. that the range goes from 1.50%-1.99%). Hence, for this reason, and the fact that all the probability associated with a specific range is allocated to the mid-point of the range and not distributed over the range, the resulting measure of skew should be interpreted with caution, particularly when it lies in the range 0.00 percentage point to 0.05 percentage point.

Considering the ECB's definition of price stability, the probability that long-term inflation may eventually be equal to or above 2% may be seen as a useful additional indicator of the risk that forecasters associate with the ECB failing to meet its objective. According to SPF respondents, there has always been some probability associated with a five-year ahead inflation outcome being above 2%. Between 2001 and 2003, this probability increased gradually from around 35% to more than 40% (see Chart 20). Since then, it fluctuated between 40% and 50%. The interpretation of this indicator, however, is not unambiguous. It captures changes in the probability mass in the upper tail of the inflation distribution, but provides no information either about the shape of the distribution or about changes in the mean (or mode) associated to it. It may nevertheless be related to the inflation risk that market participants might want to hedge when investing in inflation-linked bonds. Indeed, there appears to be some co-movement between the break-even inflation rates extracted from the yield differences between conventional government bonds and inflation-linked bonds and the probability from the SPF that long-term inflation may be equal to or above 2%. The correlation coefficient between the two series was 0.68 percentage point over the period 2002-2006.

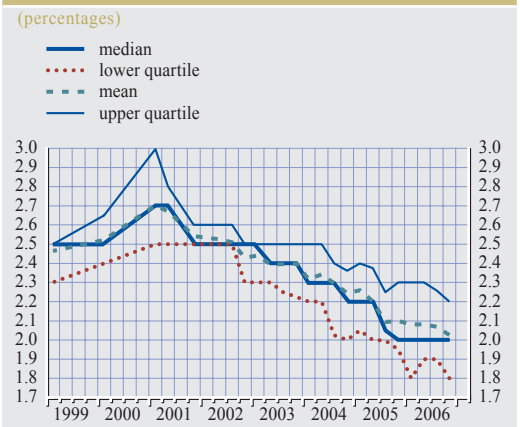
This may be because break-even inflation rates should incorporate not only the mean of market participants' inflation expectations, but also an

**Chart 21 Uncertainty surrounding SPF long-term inflation expectations**




Source: ECB calculations.

**Chart 22 SPF longer-term GDP growth expectations**




Sources: ECB and ECB calculations.

inflation risk premium (which, if positive, tends to push up break-even inflation rates) and a liquidity premium (which compensates investors for the lower liquidity of inflation-linked vis-à-vis conventional government bonds and tends to reduce break-even inflation rates).[52] Quantifying these premia is far from straightforward. It is likely that the liquidity premium charged on inflation-linked bonds has declined over time, with the growth of the euro area market for inflation-linked government debt over recent years. Consequently, it is very likely that the inflation risk premium dominates at present. This, in turn, might explain why the break-even inflation rate has mostly stayed somewhat above 2% since late 2003, whilst average SPF forecasts have remained broadly unchanged at 1.9%. Hence, given that the SPF probability that long-term inflation will be equal to or above 2% could be a proxy for the risk market participants wish to hedge, the link with the break-even inflation rate may not be surprising, given that the latter contains some elements of risk premium.

Finally, one may assess the overall uncertainty assigned to long-term inflation expectations using the uncertainty measures discussed in Section 3, and in particular, the decomposition between individual uncertainty and disagreement amongst forecasters (see Chart 21). The overall uncertainty surrounding longer-term inflation expectations has remained broadly stable over time, fluctuating around 0.6 percentage point. Interestingly, the chart also confirms the significant decline in disagreement about long-term inflation forecasts among SPF respondents which reflects the growing consensus that inflation will be below, but close to, 2%. This pattern may be linked to the clarification of the quantitative definition of price stability which occurred around the time of the review of the ECB's monetary policy strategy in 2003. By contrast, individual uncertainty appears to have increased, albeit mildly, over time. The combination of less disagreement and greater individual uncertainty means that the combined forecast uncertainty also increased over time.

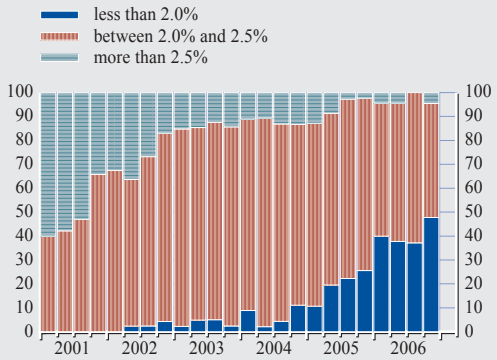
## 4.2 LONG-TERM REAL GDP GROWTH EXPECTATIONS: A VIEW ON TREND POTENTIAL OUTPUT GROWTH

Long-term real GDP growth expectations can provide indications about how trend potential output growth is perceived, including its evolution over time and the uncertainty surrounding it. Chart 22 shows that, having peaked at around 2.7% in early 2001 at the

52 For a more detailed discussion of the premia embodied in break-even inflation rates, see Kwan (2005) and ECB (2004b).

**Chart 23** Percentage of SPF respondents reporting long-term GDP growth expectations within certain ranges

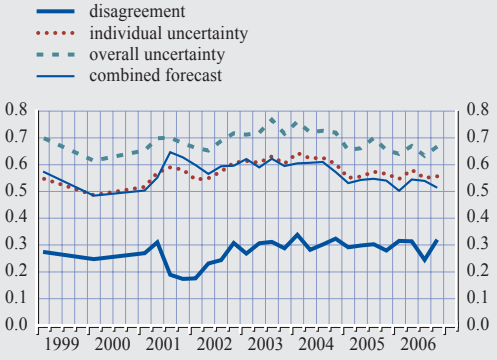
(percentages)




Source: ECB calculations.

**Chart 24** Uncertainty surrounding SPF longer-term GDP growth expectations

(percentage point)




Source: ECB calculations.

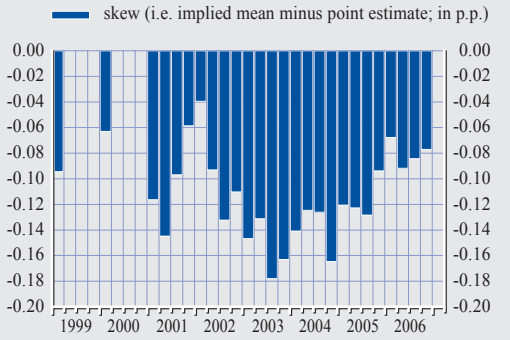
height of the new economy debate, long-term growth expectations have declined steadily, reaching 2.0% in the SPF round for the fourth quarter of 2006. Whilst part of this decline may reflect a toning down of earlier exuberant expectations, some of it is undoubtedly due to the poor growth performance of the euro area economy observed since 2001. It is also interesting to note that this increased pessimism is relatively common across SPF panel members, as indicated by the downward parallel movement of both the 25th and 75th percentiles.

Although average long-term growth expectations have been progressively revised downwards since early 2001, the frequency distribution of point forecasts reveals that most respondents continue to report long-term growth expectations in the 2.0%-2.5% range (see Chart 23). This is consistent with the ECB's estimated euro area trend potential growth (see ECB, 2005a). However, this average estimate does conceal the fact that many of the respondents who originally reported long-term growth expectations at the higher end of this range are now making forecasts at the lower end of this range. Furthermore, in the most recent rounds, only a few respondents reported long-term growth expectations above 2.5%, whilst the percentage reporting below 2.0% rose to above 10% for the first time in the SPF round for the fourth quarter of 2004 and to 48% in the round for the fourth quarter of 2006.

In addition to the decline in actual long-term growth expectations, the uncertainty about these expectations increased somewhat between early 2000 and 2004, although it has since declined again (see Chart 24). Recalling the fact that aggregate uncertainty may be decomposed into average individual uncertainty and disagreement among forecasters about the point forecasts, it appears that the profile of overall uncertainty has been driven by a combination of changes in the individual assessment of uncertainty and changes in the degree of disagreement between individuals (although the latter remained relatively low). The relatively low level of disagreement is also reflected in the small differences between the combined forecast uncertainty and the average individual uncertainty.

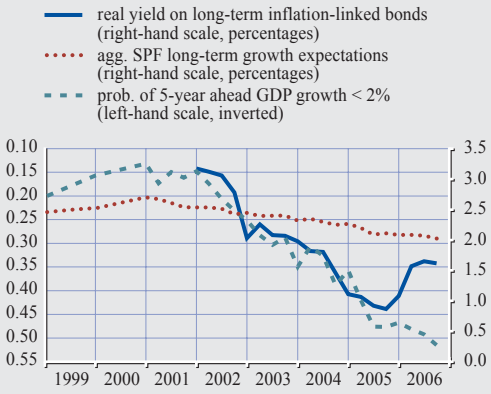
Beyond the decline in point estimates of long-term growth and the evolution of uncertainty, we also consider the balance of risks surrounding long-term growth expectations. The balance of risks to long-term growth expectations has generally been assessed to be on the downside (see Chart 25). To some extent, the fact that risks were judged to be almost permanently on the downside can be seen as consistent with the

**Chart 25 The balance of risks around long-term growth expectations**




Source: ECB calculations.
Note: There were no questions related to longer-term real GDP growth forecasts in the 1999Q2, 1999Q3, 1999Q4, 2000Q2, 2000Q3 and 2000Q4 survey rounds.

**Chart 26 Long-term growth expectations and the yield from long-term index-linked bonds**




Sources: Reuters, ECB and ECB calculations.

gradual downward revision of point estimates. Furthermore, the skew indicator was at its smallest shortly after respondents had started to revise down their long-term growth expectations in 2001. Perhaps, at that time, they felt that the downward revision of their point long-term growth expectations removed the downward risks. However, the subsequent weakness in growth may have forced them to re-evaluate the balance of risks. The qualitative comments that SPF respondents provide suggest that the overall risk assessment of longer-term GDP growth is underpinned by a number of factors. For example, in the euro area, unfavourable demographic developments in terms of lower growth in the working age population will become a major source of downward pressure on the longer-term potential output.[53] To partially offset such a negative impact, labour utilisation has to be stimulated (e.g. by increasing participation rates of women and older workers and by lowering the structural unemployment rate). However, this is largely dependant on the progress made in structural reforms in labour and product markets, and some SPF participants have expressed concerns about the lack of determination euro area countries show in implementing these necessary reforms.

We also consider the link between SPF respondents' long-term growth expectations and long-term inflation-linked bond yields. One may recall that yields on conventional bonds broadly correspond to the sum of three components: a real yield, some expected inflation and a risk premium. The real yield and the expected inflation represent the two forms of compensation investors require: a real return on the investment and compensation for the purchasing power lost as a result of inflation. In contrast to conventional bonds, the yield of inflation-linked bonds does not include expected inflation and only includes the "real" component plus an associated real risk premium, since the par value and the coupon payments automatically adjust to changes in inflation. Although influenced by a host of factors, the real yield is determined primarily by the economy's overall demand and supply of capital, which tends to follow the economy's overall rate of growth.

Indeed, over the period for which data are available (i.e. since early 2002), there has been a broad co-movement between SPF long-term growth expectations and the yield on inflation-linked bonds, although there has been some decoupling in 2006. This broad co-movement could reflect

53 See ECB (2005a).

declining long-term growth expectations from both professional forecasters and market participants. However, the scale of movement has been somewhat different. On the one hand, SPF expectations have declined (by 0.5 percentage point) from around 2.5% at the beginning of 2002 to 2.0%, whilst yields on inflation-linked bond yields declined (by 2.3 percentage point) from around 3.2% to 0.9% at their lowest – see Chart 26. Clearly, other factors may also impact on yields from long-term inflation-linked bonds, such as technical market factors (e.g. the degree of liquidity and demand in the market), the premium investors are willing to pay to be hedged against inflation risk, or unexplained factors that have fuelled the low-bond yield conundrum described by Alan Greenspan, former Chairman of the Board of Governors of the Federal Reserve.[54] Interestingly, according to SPF respondents, there has been a downward movement in both the real yield on long-term inflation-linked bonds and the probability of growth being below 2%. This may indicate that the long-term bond yields also reflect, at least in part, the markets' perceptions of risks to long-term growth.

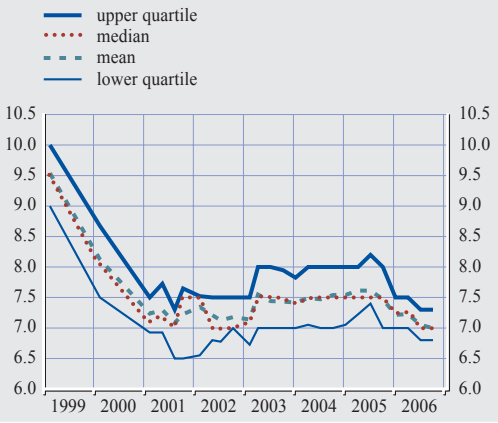
## 4.3 LONG-TERM UNEMPLOYMENT EXPECTATIONS

SPF respondents also provide long-term unemployment expectations. Following a period of declining expectations from 1999 to early 2001, long-term unemployment expectations broadly flattened out at a level of around 7.5%, before declining more recently to around 7.0% (see Chart 27). This profile has been largely shared across the SPF panel, as witnessed by the equivalent profiles of the upper and lower quartiles.

Considering Chart 28, it is clear that the profile of long-term unemployment expectations has been largely driven by revisions to the short-term unemployment outlook.[55] There seems to be a clear, and almost one-to-one, average relationship between revisions to the short-term outlook and those to the longer-term outlook. This reflects the high degree of persistence in the unemployment data and also suggests that SPF respondents perceive a high degree of hysteresis[56], although on some occasions the change in the short-term expectation has moved in the opposite direction to the long-term expectation (i.e. when the observations are in the red quadrants of Chart 28). It has also been

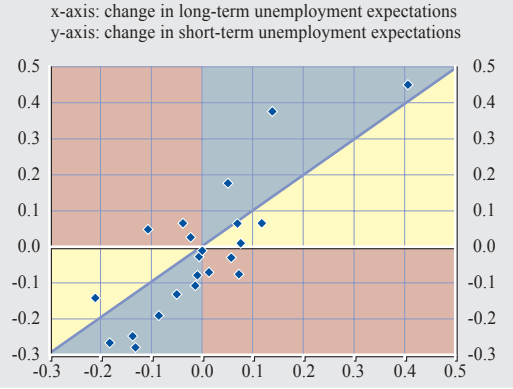
54 In a speech before the US Congress on 16 February 2005.
55 It was also considered whether there was a correlation between changes to long-term GDP and long-term unemployment expectations. However, while there appeared to be a negative relationship, it was not statistically significant.
56 If there was no hysteresis in unemployment, one could expect revisions to longer-term unemployment expectations to be uncorrelated with short-term developments.

**Chart 27 Evolution of long-term unemployment expectations**



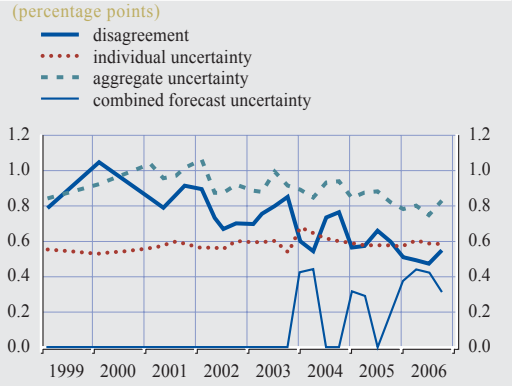

Sources: ECB and ECB calculations.

**Chart 28 Correlation between revisions to the short-term unemployment outlook and the long-term unemployment outlook**



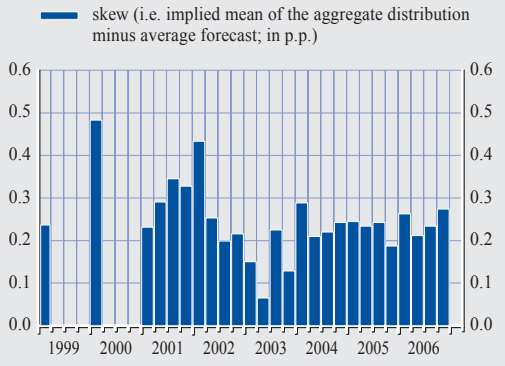

Source: ECB calculations.

**Chart 29 Uncertainty surrounding SPF longer-term unemployment rate forecasts**




Source: ECB calculations.
Note: There were no questions related to longer-term unemployment forecasts in the 1999 Q2, 1999 Q3, 1999 Q4, 2000 Q2, 2000 Q3 and 2000 Q4 survey rounds.

**Chart 30 The balance of risks around long-term unemployment**




Source: ECB calculations.

the case on a number of occasions that the revisions to the long-term expectations have been larger than those to the short-term expectations (i.e. when the observations are in the yellow quadrants). The blue quadrants indicate when the revision of the short-term expectation is larger than that of the long-term expectation. The high degree of persistence, however, is amplified by the substantial data revisions that have affected the unemployment statistics and, consequently, SPF forecasts over all horizons (see Section 2.1).

Another noteworthy feature of the longer-term SPF unemployment expectations is that the level of disagreement between forecasters (as measured by the standard deviation of point forecasts) has generally been higher than their individual assessment of the uncertainty surrounding their point forecasts (as measured by the average standard deviation of the individual distributions). This can be seen clearly in Chart 29. Up until mid-2003, individual forecasters were relatively confident in their own assessment of the uncertainty surrounding their forecast, particularly when assessed against the high degree of disagreement among forecasters about the longer-term unemployment forecast. Since mid-2003, there has been a further small decline in the level of disagreement, while the individual assessment of uncertainty has remained broadly stable. Unlike inflation or real GDP growth which can be anchored by monetary policy or potential output expectations, long-term unemployment expectations may differ substantially from one forecaster to the next according to how different information is weighted in forming expectations or, alternatively, to the type of model used to assess its likely future evolution. This reflects the fact that, in the longer term, unemployment rates are determined by underlying labour market features. Therefore, participants' assessments of the likelihood and magnitude of structural reforms play an important role in terms of both the level of their longer-term unemployment rate forecasts and the risks surrounding their forecasts.

Finally, the balance of risks assigned to longer-term unemployment expectations has generally been assessed to be on the upside. Chart 30 shows the perceived balance of risks, proxied by the difference between the mean of the aggregate distribution and the average of the point forecasts. This continuous perception of an upward balance of risks, despite revisions to the point forecast, may be due to respondents' scepticism that required structural reforms will take place in the labour market. Indeed, the factor most frequently cited by SPF respondents as having an impact on their long-term unemployment expectations is

the prospect (or lack) of structural reforms in the labour market.

## 5 CONCLUDING REMARKS AND FUTURE SPF RESEARCH

Since its inception, the ECB SPF has led to the accumulation of a valuable set of data on expectations about inflation, real GDP growth and unemployment in the euro area from a large panel of professional forecasters. Whilst the number of observations is still small, a number of interesting conclusions can be drawn from this first assessment of the SPF results. First, in terms of actual forecast performance, the SPF has been broadly in line with other comparable surveys of forecasters (such as Consensus Economics). More specifically, this implies that, over the period under consideration (first quarter of 1999 to the fourth quarter of 2006), inflation has tended to be underpredicted to a significant extent, whilst growth has tended to be overpredicted, albeit to a lesser extent. Interestingly, there is little systematic difference between the performance of panel members irrespective of whether they are from the financial or non-financial sector or of the country in which they are based. Second, the SPF has also enabled us to measure forecasters' assessment of the uncertainty surrounding their point forecasts. The analysis suggests that, in line with their point forecast errors, SPF participants may not have accurately assessed the uncertainty surrounding their forecasts. For example, the one-year ahead inflation outcome has nearly always been in the upper half of the aggregate probability distribution. As regards the level of uncertainty, the pattern of short-term uncertainty perceived by SPF participants has been broadly stable for both inflation and unemployment since the start of the SPF. In some contrast, the SPF indicates a higher degree of variability for short-term GDP growth, especially during the period 2002-03. Lastly, the SPF also enables us to consider different aspects of participants' longer-term expectations. One key feature of this long-term picture is the stable anchoring of long-term inflation expectations at a level in line with the definition of price stability adopted by the ECB Governing Council, which may be read as a sign of the ECB's credibility.

This study has provided an initial assessment of the SPF data on euro area macroeconomic expectations. However, the main questions that have been addressed will clearly need to be looked into again once more surveys have been carried out and more data have been collected. In particular, our analysis has been conducted over a period that arguably does not yet constitute one full business cycle if measured from the peak in euro area GDP growth in mid-2000. A more reliable assessment of the performance of the SPF forecasters would be based on not just one but several complete euro area business cycles. This would be particularly justified given the uncertainties about the structure and functioning of the euro area economy and that SPF participants have had to learn about the shocks impacting on the economy, as well as the mechanisms by which they are propagated to inflation, growth and unemployment. A main focus of future research should therefore be to re-examine – with the benefit of greater hindsight – the predictive accuracy of the SPF macroeconomic expectations and density forecasts and to check further the robustness of the main results established here.

One important additional area of research that has not been pursued in this paper is to attempt to understand more fully the economic forces determining macroeconomic expectations and, in this regard, to assess more formally the extent to which they deviate from the full-information or rational paradigm. In assessing euro area macroeconomic developments, the SPF panel members face a number of uncertainties, including model estimation and inference problems. This may give rise to deviations from the knowledge that is imputed to agents under the more extreme versions of the Rational Expectations Hypothesis (REH – see Sargent, 1993). For the US economy, a number of such studies of the REH have already

been undertaken using panels of professional forecasters similar to the ECB SPF.[57] In general, such studies have not been conclusive, although the balance of evidence appears to sway slightly in favour of less than full rationality. Clearly, while the still relatively small time-dimension of the SPF sample means that such work must be put on hold for now, it will be of great interest to consider the "efficiency" of the SPF expectations. In particular, future work should focus on trying to assess the extent to which panel members have (or have not) optimally made use of the available information, including their own past forecast errors, when forming their expectations.[58] As the SPF panel members are "professional" forecasters, who may have access to a lot of relevant macroeconomic data and models, it will also be interesting to compare the efficiency of their expectations with that of other agents in the economy, including households. Recent research using US data has shown some surprising results in this regard, in particular that household expectations compare reasonably favourably with those of professional forecasters (see, for example, Thomas, 1999).

Another direction for future work would be to examine more fully the heterogeneity of the SPF panel. For example, do some forecasters appear to be less biased, more rational or learn more quickly about likely future economic events than others? In the area of forecast uncertainty, this study has focused only on the evaluation of the aggregate density forecasts. It would therefore be of interest to conduct an evaluation of the density forecasts of individual panel members in order to assess whether some members of the panel are better and more able to assess accurately macroeconomic risks than others.

A final direction for future work on the SPF data could be to consider their introduction into macroeconomic models in order to capture the impact of forward-looking behaviour on economic variables. As discussed in Carroll (2001), macroeconomists have devoted little effort to modelling actual empirical expectations data in recent decades; instead they have tended to assume all agents are rational. However, the increasing availability of datasets such as the SPF creates the opportunity to understand and model much more realistically the mechanisms whereby expectations are formed and how they may impact on the macroeconomy. For example, a number of studies have been undertaken using survey data that test for a forward-looking component in the inflation process. There are numerous other examples of the relevance of inflation expectations from a central bank's perspective. Most notably, inflation expectations can be viewed as a determinant of money demand and they are likely to play an important role in the monetary transmission mechanism. In the recent literature on monetary policy rules, inflation expectations have also been posited to be among the key arguments in a central bank's reaction function (Taylor, 1999). Of course, the role of expectations in economic modelling extends far beyond inflation and monetary policy; hence, the SPF data could also eventually play a role in helping to model important real variables, such as consumption and investment.

57 See, for example, Thomas (1999), Brown and Maital (1981) and Baghestani and Kianian (1993). Keane and Runkle (1990) do not reject rationality, while Zarnowitz (1985) finds that forecasts of other variables, except prices, tend to be rational. See Croushore (1997) for an overview of results investigating the Livingston Survey in the United States.

58 For a further discussion of efficiency of expectations, see Thomas (1999) and some of the references cited therein.

## ANNEX I

### IS THERE BIAS IN THE SPF FORECAST RESULTS? EVIDENCE FROM A PANEL DATA ANALYSIS

Section 2.2 presented some tests for unbiasedness performed on aggregate or average survey responses. Some authors have criticised the use of average responses to perform these tests since the aggregation of survey responses may introduce at least two types of bias. First, Figlewski and Wachtel (1983) observed that the use of average survey responses introduces a specification bias. Indeed, if all forecasters were rational (i.e. they were to all pass the bias test regression), it would mean that the differences between individual forecasts would only be the result of differences in private information sets. However, because each forecaster cannot use the private information of other individuals, Figlewski and Wachtel show that the average forecast is correlated with the average forecast error (which contains the private information variables) and therefore is not rational. This correlation would introduce a substantial bias into the parameters of the test regression. Second, average forecasts may mask systematic individual biases and thus, lead to a false acceptance of unbiasedness (Keane and Runkle, 1990).

The aggregation bias is indeed quite significant and can be simply illustrated by running three types of regression and comparing the resulting coefficients with those presented in Section 2.2.[59] Regression 1 uses survey averages, as in Section 2.2, regression 2 involves doing a separate regression for each individual SPF forecaster and taking the mean of the estimated coefficients, and regression 3 uses pooled (combined time series, cross section) data. Table 10 reports the estimated α and β coefficients resulting from these regressions, using the one-year ahead forecast of inflation.

Table 10 shows that β is closer to unity when using SPF average survey responses, so that the aggregation of SPF replies actually increases the likelihood of accepting the null hypothesis of forecast unbiasedness.[60] The upward bias in the β estimator of the regression using survey averages appears to be quite severe and justifies the attempt to use pooled data in a cross-section regression to accurately test for bias in SPF results.

Using individual forecaster level data could therefore overcome aggregation problems by modelling differences in behaviour across each individual forecaster (see also Lahiri and Liu, 2004). In this Annex, we run similar bias tests to those of Section 2.2 using the following equation:

$$A_t = \alpha_i + \beta_i F_{it} + \varepsilon_{it}$$

where $F_{it}$ is the forecast of forecaster $i$ made for time $t$.

59 In what follows, we used a balanced panel (i.e. for each variable/horizon, we only use results from those who have replied in each survey round for that variable/horizon). As only a very small number of panellists have replied in every round, to ensure that a sufficient number of responses were available, a simple linear interpolation was made (provided that the gap between responses was no more than two rounds). The resulting balanced panel contains approximately 30 respondents for the one-year ahead rolling horizon and just over 20 for the two-year ahead rolling horizon.

60 A similar aggregation bias was found by Figlewski and Wachtel (1981) in a study of the Livingston survey data on inflation expectations and by Urich and Wachtel (1984) in a study on money supply expectations.

**Table 10 Regression test for bias of one-year ahead forecast of inflation**

| Estimated coefficients | α | β |
|---|---|---|
| 1. SPF average survey responses | 1.03 | 0.68 |
| 2. Individual means | 1.85 | 0.20 |
| 3. Pooled data | 1.95 | 0.14 |

Source: ECB calculations.

**Table 11 Results from bias test using averages of balanced SPF panel data**

| Variable | α | β | $R^2$ | F-test (p-value) | μ (p-value) |
|---|---|---|---|---|---|
| Inflation one year ahead | 1.03 (0.40) [1] | 0.68 (0.24) | 0.12 | **0.000** | **0.000** |
| Inflation two years ahead | 3.51 (0.77) [1] | -0.73 (0.44) [1] | 0.06 | **0.000** | **0.000** |
| GDP one year ahead | -0.35 (0.63) | 1.08 (0.36) | 0.28 | 0.691 | 0.603 |
| GDP two years ahead | 4.05 (1.47) [1] | -0.95 (0.60) [1] | 0.07 | **0.011** | **0.003** |
| Unemployment one year ahead | 7.71 (1.07) [1] | 0.08 (0.11) [1] | 0.01 | **0.000** | 0.106 |
| Unemployment two years ahead | 11.76 (0.83) [1] | -0.40 (0.09) [1] | 0.38 | **0.000** | 0.933 |

Source: ECB calculations.
Note: Standard errors are in brackets (corrected using the Newey-West procedure).
1) Indicates that the coefficient is significantly different from the null hypothesis (i.e. α = 0 or β = 1) at the 10% level or lower. The p-value is from the F-test. Bold indicates rejection of the unbiasedness hypothesis at the 10% level or lower.

Since the dependent variable ($A_t$), i.e. the observed outcome of the forecast variable, is the same for all individual forecasters, forecast errors are likely to be highly correlated across forecasters due to unforeseen aggregate shocks to the economy, and thus common to all survey participants. Moreover, each individual forecast error ($\varepsilon_{it}$) suffers from autocorrelation due to the overlapping forecast horizons noted in Section 2.2. Hence, the estimation was performed using feasible GLS, allowing for autoregressive and heteroskedastic residuals.

As highlighted above, the panel data analysis was carried out using a balanced panel, constructed where necessary by linear interpolation of missing forecasts. By doing so, we are imposing a specific forecast generating process on the individual forecaster that may influence fully-fledged tests for rational expectations of the SPF forecasts. However, as our aim here is to measure only the statistical significance of the forecast bias, we implement this simplifying method.

As a first step, we check whether the average results of the constructed balanced panel are comparable with those of the full sample presented in Section 2.2. Table 11 reports the results of the bias tests performed using the average results of the balanced panel, and shows that they are indeed broadly similar to those of the unbalanced panel, as presented in Section 2.2. More precisely, the test results indicate a significant bias in the inflation forecasts over all horizons, as well as in the two-year ahead real GDP growth forecasts. The bias does not appear to be significant for unemployment forecasts or for one-year ahead real GDP growth forecasts.

**Table 12 Results from bias tests using pooled data of the SPF balanced panel**

| Variable | α | β | μ (p-value) |
|---|---|---|---|
| Inflation one year ahead | 1.94 (0.05) [1] | 0.14 (0.03) [1] | **0.000** |
| Inflation two years ahead | 2.36 (0.07) [1] | -0.09 (0.04) [1] | **0.000** |
| GDP one year ahead | 1.93 (0.11) [1] | 0.14 (0.03) [1] | 0.309 |
| GDP two years ahead | 2.12 (0.17) [1] | -0.07 (0.06) [1] | **0.000** |
| Unemployment one year ahead | 7.13 (0.16) [1] | 0.13 (0.02) [1] | **0.000** |
| Unemployment two years ahead | 8.32 (0.14) [1] | -0.03 (0.01) [1] | **0.000** |

Source: ECB calculations.
Note: Standard errors are in brackets.
1) Indicates that the coefficient is significantly different from the null hypothesis (i.e. α = 0 or β = 1) at the 10% level or lower. Bold indicates rejection of the unbiasedness hypothesis at the 10% level or better.

Table 12 reports the corresponding results of the bias tests using the pooled data of that same balanced panel. Clearly, the probabilities of SPF results being unbiased are now equal to zero for all variables, with the exception of GDP one year ahead. Furthermore, $\alpha > 0$ and $\beta < 1$ in all cases. This suggests that aggregating SPF results introduces a significant bias into the test results by masking systematic individual biases. However, it should be borne in mind that forecast bias results for the SPF are likely to be sensitive to the small sample size and should thus be considered with caution.

## ANNEX 2

### A CLOSER LOOK AT INDIVIDUAL FORECASTER QUALITY

This Annex highlights a number of alternative indicators for assessing the "quality" of SPF forecasters and provides further and more detailed information on the performance of each individual SPF forecaster. Presented in the table below, these indicators provide a more comprehensive picture of the behaviour of individual panel members. They include the response rate, forecast accuracy and the degree to which a specific forecaster lags or leads the rest of the panel.

*Response rates*: the table shows (in the second, third and fourth columns respectively) the number of responses from each forecaster in each SPF round and more specifically the number of responses to the questions on their point forecasts and the underlying probability distributions as a percentage of the total number of possible responses. While no forecaster has an absolute 100% response rate, quite a number have responded consistently to most of the questions asked. For example, 11 have responded at least 90% of the time to the questions regarding their point forecasts, and 32 have responded at least 75% of the time. The equivalent figures for the probability distributions are 10 and 27.

*Accuracy*: in terms of forecast error statistics, the table presents the rankings with regard to the RMSE statistics across the three different variables (inflation, growth and unemployment) and two horizons (one-year and two-year ahead) in order to consider whether any forecaster has performed significantly above or below the average (see the fifth and sixth columns).[61] Taking the upper and lower quartiles as the thresholds, based on the experience to date, seven forecasters (numbers 22, 23, 24, 60, 80, 93 and 96) have better-than-average RMSE statistics (in the upper quartile) for at least four of the different six combinations of variables/horizons considered. However, it should be noted that none of these forecasters provided answers to the rolling horizon questions during the earlier rounds of the SPF when errors were relatively high. There were also five forecasters (numbers 3, 19, 28, 71 and 87) who have worse-than-average RMSE statistics (in the lower quartile) for four or more of the different variables/horizons for which they provided forecasts. Of these, one forecaster withdrew from the panel after the round for the fourth quarter of 2003. Overall, even though a small number of forecasters perform above or below average across a number of the variables and horizons, none do so for all variables and all horizons. Further evidence and experience would therefore be required before we are able to systematically identify good performers.

*Leading or lagging the group*: The seventh and eighth columns of the table show the extent to which changes in individual forecasts lead or lag changes in the average SPF forecast. Forecasters systematically lagging the average could be slow to adjust their views to reflect new information or they could be merely updating their forecasts with the consensus. On the other hand, if forecasters are consistently ahead of or at least contemporaneous with the average SPF forecast, it would suggest that their forecasts are indeed being updated to reflect new information.[62] In any case, the results show that no forecaster shows a positive and significant contemporaneous or leading correlation with changes in the average forecast

61 Only forecasters who have responded at least ten times for a specific variable/horizon combination are taken into account in the ranking for that specific variable/horizon combination.

62 However, a number of caveats to this analysis should be highlighted. First, given that the majority of respondents do not reply all the time, the change in their forecasts is put at zero if they did not reply in the previous round. Second, perhaps it is not the correlation with the average SPF forecast but the correlation with the actual value that matters. However, given that the actual outcome is not known until approximately 12 months later and that many things could happen in the meantime, the correlation with changes in the average forecast is considered, as it should contain most of the information that has become available since the last round. Lastly, persistence in changes in the forecasts, arising possibly either from the inertia of some forecasters or from persistence in shocks, suggests these correlations should be interpreted with caution. This is particularly the case for the unemployment forecasts (both one and two years ahead) as the changes in the average forecast are significantly correlated with changes in the previous round. Thus, a positive correlation with lagged changes may not only indicate that a forecaster is following the consensus, but also that shocks may be correlated.

for all six variables/horizons,[63] although a couple show a positive and significant (more than 0.4 percentage point) correlation for five of the horizons/variables (numbers 3 and 4). Similarly, no forecasters consistently lag changes in the average SPF forecast, although one does lag for three variables/horizons (number 41). Lastly, it should be highlighted that no forecaster was a "contrarian" forecaster, as none had a negative and significant contemporaneous correlation with changes in the average forecast.

To sum up, considering the different aspects of forecaster quality (response rate, forecast accuracy, etc.) and bearing in mind the short sample for which actual outcomes can be compared against forecasts, it is difficult to highlight conclusively and systematically "good" or "bad" forecasters. This is especially the case when one considers that the period since 1999 has been characterised by a number of shocks, in particular to oil prices. Given that these, to a large extent could not be forecast, a forecaster that has performed particularly well over the sample period might actually have performed relatively poorly in the absence of such shocks and vice versa.

63 The results here are shown for six variants: the one and two-year ahead rolling horizons for inflation, GDP growth and unemployment.

## Table 13 Alternative indicators of forecaster quality

| Forecaster | Response rate – rounds [1] (%) | Response rate – point forecasts [2] (%) | Response rate – probability distribution [3] (%) | Forecast ranking – top quartile [4] | Forecast ranking – bottom quartile [5] | Leading/ contem-poraneous with aggregate [6] | Lagging aggregate [7] |
|---|---|---|---|---|---|---|---|
| SPF | | | | 0/6 | 0/6 | 6/6 | 2/6 |
| 1 | 81 | 78 | 79 | 0/6 | 0/6 | 4/6 | 2/6 |
| 2 | 69 | 64 | 66 | 3/6 | 0/6 | 2/6 | 1/6 |
| 3 | 72 | 64 | 68 | 0/6 | 4/6 | 5/6 | 2/6 |
| 4 | 91 | 87 | 85 | 2/6 | 1/6 | 5/6 | 0/6 |
| 5 | 88 | 84 | 87 | 1/6 | 0/6 | 2/5 | 1/5 |
| ~~6~~ | ~~6~~ | ~~6~~ | ~~6~~ | | | | |
| 7 | 94 | 75 | 0 | 0/6 | 2/6 | 4/5 | 1/5 |
| ~~8~~ | ~~0~~ | ~~0~~ | ~~0~~ | | | | |
| 9 | 81 | 80 | 80 | 0/6 | 3/6 | 4/6 | 2/6 |
| 10 | 75 | 67 | 67 | 0/6 | 1/6 | 2/6 | 0/6 |
| 11 | 91 | 84 | 81 | 0/6 | 3/6 | 3/5 | 0/5 |
| ~~12~~ | ~~0~~ | ~~0~~ | ~~0~~ | | | | |
| ~~13~~ | ~~3~~ | ~~3~~ | ~~3~~ | | | | |
| 14 | 94 | 89 | 69 | 0/6 | 2/6 | 3/6 | 1/6 |
| 15 | 81 | 82 | 82 | 2/6 | 0/6 | 1/6 | 0/6 |
| 16 | 84 | 83 | 83 | 2/6 | 0/6 | 2/6 | 0/6 |
| 17 | 81 | 77 | 71 | 2/6 | 3/6 | 0/5 | 0/5 |
| 18 | 34 | 30 | 17 | | | | |
| 19 | 78 | 78 | 78 | 0/6 | 4/6 | 3/6 | 1/6 |
| 20 | 81 | 78 | 81 | 1/5 | 0/5 | 2/5 | 0/5 |
| ~~21~~ | ~~0~~ | ~~0~~ | ~~0~~ | | | | |
| 22 | 78 | 79 | 79 | 5/6 | 1/6 | 2/6 | 1/6 |
| 23 | 78 | 77 | 77 | 4/6 | 0/6 | 4/6 | 1/6 |
| 24 | 91 | 91 | 91 | 4/6 | 0/6 | 4/5 | 0/5 |
| ~~25~~ | ~~0~~ | ~~0~~ | ~~0~~ | | | | |
| 26 | 100 | 95 | 100 | 0/6 | 0/6 | 3/5 | 2/5 |
| ~~27~~ | ~~0~~ | ~~0~~ | ~~0~~ | | | | |
| 28 | 78 | 68 | 69 | 0/6 | 5/6 | 4/6 | 0/6 |
| 29 | 78 | 49 | 49 | 1/3 | 1/3 | 1/5 | 0/5 |
| 30 | 81 | 48 | 48 | 0/1 | 0/1 | | |
| 31 | 91 | 89 | 56 | 1/6 | 1/6 | 3/6 | 0/6 |
| 32 | 94 | 66 | 67 | 2/6 | 0/6 | 3/6 | 2/6 |
| 33 | 100 | 97 | 99 | 2/6 | 2/6 | 0/5 | 0/5 |
| ~~34~~ | ~~19~~ | ~~14~~ | ~~15~~ | | | | |
| 35 | 97 | 45 | 44 | 2/3 | 0/3 | | |
| 36 | 84 | 83 | 83 | 1/6 | 2/6 | 3/6 | 0/6 |
| 37 | 97 | 97 | 97 | 2/6 | 0/6 | 4/6 | 0/6 |
| 38 | 69 | 60 | 60 | 1/6 | 2/6 | 4/6 | 0/6 |
| 39 | 94 | 93 | 93 | 1/6 | 1/6 | 4/5 | 0/5 |
| 40 | 53 | 28 | 28 | | | | |
| 41 | 97 | 97 | 56 | 3/6 | 1/6 | 3/6 | 3/6 |
| 42 | 88 | 83 | 84 | 0/6 | 1/6 | 1/5 | 0/5 |
| ~~43~~ | ~~9~~ | ~~7~~ | ~~7~~ | | | | |
| ~~44~~ | ~~6~~ | ~~3~~ | ~~3~~ | | | | |
| 45 | 91 | 38 | 34 | | | | |
| ~~46~~ | ~~28~~ | ~~24~~ | ~~24~~ | | | | |
| 47 | 88 | 50 | 46 | 1/3 | 0/3 | | |
| 48 | 84 | 55 | 44 | 3/6 | 2/6 | 1/6 | 0/6 |
| 49 | 38 | 32 | 27 | 1/3 | 1/3 | 0/4 | 1/4 |
| 50 | 59 | 49 | 49 | 1/6 | 3/6 | 1/5 | 0/5 |
| ~~51~~ | ~~0~~ | ~~0~~ | ~~0~~ | | | | |
| 52 | 69 | 67 | 68 | 2/6 | 0/6 | 4/6 | 0/6 |
| ~~53~~ | ~~66~~ | ~~64~~ | ~~63~~ | ~~0/6~~ | ~~1/6~~ | ~~4/6~~ | ~~1/6~~ |
| 54 | 97 | 95 | 94 | 2/6 | 1/6 | 3/5 | 0/5 |
| ~~55~~ | ~~28~~ | ~~24~~ | ~~23~~ | | | | |
| 56 | 91 | 81 | 82 | 2/6 | 2/6 | 4/6 | 1/6 |
| 57 | 91 | 46 | 20 | 0/1 | 0/1 | | |
| ~~58~~ | ~~3~~ | ~~2~~ | ~~1~~ | | | | |
| 59 | 88 | 85 | 80 | 0/6 | 1/6 | 3/6 | 0/6 |
| 60 | 72 | 72 | 72 | 4/6 | 0/6 | 1/6 | 1/6 |
| 61 | 75 | 64 | 64 | 1/6 | 1/6 | 3/5 | 0/5 |
| ~~62~~ | ~~38~~ | ~~17~~ | ~~17~~ | | | | |

## Table 13 Alternative indicators of forecaster quality (cont'd)

| Forecaster | Response rate – rounds [1] (%) | Response rate – point forecasts [2] (%) | Response rate – probability distribution [3] (%) | Forecast ranking – top quartile [4] | Forecast ranking – bottom quartile [5] | Leading/ contem-poraneous with aggregate [6] | Lagging aggregate [7] |
|---|---|---|---|---|---|---|---|
| SPF | | | | 0/6 | 0/6 | 6/6 | 2/6 |
| 63 | 84 | 59 | 49 | 1/2 | 1/2 | 0/4 | 0/4 |
| ~~64~~ | ~~41~~ | ~~31~~ | ~~28~~ | ~~0/1~~ | ~~0/1~~ | | |
| 65 | 31 | 26 | 27 | | | | |
| ~~66~~ | ~~16~~ | ~~14~~ | ~~9~~ | | | | |
| 67 | 72 | 53 | 43 | 1/6 | 3/6 | 1/5 | 2/5 |
| 68 | 59 | 44 | 43 | 0/2 | 0/2 | | |
| ~~69~~ | ~~0~~ | ~~0~~ | ~~0~~ | | | | |
| 70 | 81 | 73 | 73 | 3/6 | 0/6 | 1/6 | 0/6 |
| ~~71~~ | ~~50~~ | ~~49~~ | ~~49~~ | ~~0/6~~ | ~~4/6~~ | ~~2/6~~ | ~~2/6~~ |
| 72 | 63 | 61 | 60 | 0/5 | 3/5 | 4/6 | 0/6 |
| 73 | 100 | 54 | 54 | 0/3 | 1/3 | | |
| ~~74~~ | ~~0~~ | ~~0~~ | ~~0~~ | | | | |
| ~~75~~ | ~~0~~ | ~~0~~ | ~~0~~ | | | | |
| 76 | 88 | 74 | 68 | 1/6 | 1/6 | 3/5 | 0/5 |
| ~~77~~ | ~~0~~ | ~~0~~ | ~~0~~ | | | | |
| ~~78~~ | ~~0~~ | ~~0~~ | ~~0~~ | | | | |
| ~~79~~ | ~~0~~ | ~~0~~ | ~~0~~ | | | | |
| 80 | 72 | 72 | 32 | 5/6 | 0/6 | 3/5 | 1/5 |
| ~~81~~ | ~~0~~ | ~~0~~ | ~~0~~ | | | | |
| 82 | 53 | 44 | 20 | 2/3 | 1/3 | 0/4 | 0/4 |
| ~~83~~ | ~~0~~ | ~~0~~ | ~~0~~ | | | | |
| 84 | 63 | 48 | 49 | 1/2 | 0/2 | | |
| 85 | 100 | 70 | 61 | 0/4 | 3/4 | | |
| ~~86~~ | ~~3~~ | ~~3~~ | ~~3~~ | | | | |
| ~~87~~ | ~~59~~ | ~~57~~ | ~~57~~ | ~~0/6~~ | ~~4/6~~ | ~~3/6~~ | ~~1/6~~ |
| 88 | 44 | 42 | 42 | 0/3 | 2/3 | 3/6 | 0/6 |
| 89 | 97 | 98 | 98 | 1/6 | 1/6 | 3/6 | 0/6 |
| 90 | 97 | 93 | 92 | 0/6 | 1/6 | 4/6 | 2/6 |
| 91 | 91 | 78 | 81 | 2/6 | 0/6 | 2/6 | 2/6 |
| 92 | 91 | 73 | 65 | 1/6 | 1/6 | 3/6 | 0/6 |
| 93 | 97 | 88 | 89 | 5/6 | 0/6 | 3/6 | 0/6 |
| 94 | 97 | 95 | 97 | 0/6 | 2/6 | 3/6 | 0/6 |
| 95 | 97 | 97 | 97 | 0/6 | 1/6 | 3/6 | 2/6 |
| 96 | 84 | 64 | 63 | 5/5 | 0/5 | 2/6 | 0/6 |
| 97 | 38 | 38 | 39 | | | 0/5 | 0/5 |
| 98 | 19 | 20 | 15 | | | | |
| 99 | 25 | 27 | 27 | | | | |
| 100 | 13 | 11 | 11 | | | | |

Notes: Strikethrough and darker background denotes that the forecasters are no longer active members of the panel, i.e. they have not replied in any of the last eight rounds.
1) "Response rate – rounds" denotes the percentage of survey rounds to which the forecaster has responded.
2) "Response rate – point forecasts" denotes the percentage of questions on their point forecasts to which the forecaster has responded. For example, in most rounds respondents are asked to provide 15 point forecasts (three variables for three calendar-year horizons and two rolling horizons)
3) "Response rate – probability distributions" denotes the percentage of questions on the probability distributions surrounding their point forecasts to which the forecaster has responded. For example, in most rounds respondents are asked to provide 15 probability distributions (three variables for three calendar-year horizons and two rolling horizons)
4) "Forecast ranking – top quartile" denotes the number of times the forecaster ranks in the top quartile in terms of RMSE for the one and two-year ahead rolling horizons for the three variables. Note: forecasters are only ranked if they have replied at least ten times. Hence "2/4" denotes that a forecaster has only responded a minimum of ten times for four out the possible six forecasts (two horizons and three variables) and was in the top quartile for two of these.
5) "Forecast ranking – bottom quartile" as above, but referring to the bottom quartile.
6) "Leading/contemporaneous with aggregate" denotes whether changes in the individual's forecasts were either leading or contemporaneously correlated with changes in the aggregate forecast. Note: the correlation is only calculated if a forecaster has replied at least ten times. Hence "2/4" denotes that a forecaster has only responded a minimum of ten times for four out the possible six forecasts (two horizons and three variables) and the correlation coefficient was significant, either leading or contemporaneously, for two of these.
7) "Lagging aggregate" denotes whether changes in the individual's forecasts were lagging changes in the aggregate forecast. Note: the correlation is only calculated if a forecaster has replied at least ten times. Hence, "2/4" denotes that a forecaster has only responded a minimum of ten times for four out the possible six forecasts (two horizons and three variables) and the correlation coefficient (for lagged changes) was significant for two of these.

## ANNEX 3

### THE PROBABILITY FORECASTS FROM THE ECB SPF: A PRELIMINARY ASSESSMENT

In addition to providing point estimates, SPF respondents provide information on their assessment of the uncertainty surrounding these point estimates. As was the case with the point forecasts, it may also be insightful to evaluate the accuracy and reliability of these assessments of uncertainty. In this Annex, we apply a method, known as the "probability integral transform", proposed by Diebold, Gunther and Tay (1998) to assess the information on forecast uncertainty that can be obtained from the ECB SPF. After briefly describing the intuition behind this approach to evaluating distribution forecasts, the forecast densities (or probability distributions) for each of the three variables (inflation, unemployment and GDP) and over two horizons (one-year and two-year ahead) are evaluated. The sample period, which ranges from the first quarter of 1999 to the fourth quarter of 2006, is relatively short and thus provides more of an illustration of an evaluation framework that can be maintained in order to assess the SPF distributions, as such series accumulate over time. In particular, a considerably larger sample would be required in order to carry out formal tests on the optimality of the SPF forecast densities. This first analysis is therefore more heuristic, based on graphical analysis and simple statistical tests.

### THE PROBABILITY INTEGRAL TRANSFORM

At first glance, the problem of evaluating density forecasts seems very challenging. In the case of point forecasts, one can observe the corresponding realisation which serves as a benchmark for predictive evaluation. However, in the case of density forecasts, there appears to be no such observable benchmark that could facilitate a test of how closely reported densities correspond to the unobservable true density of the variable under consideration. Moreover, the true density underlying the forecast variable may exhibit structural changes, thereby further complicating the evaluation process.

Drawing on a related literature evaluating interval forecasts, recent advances in forecast evaluation have begun to address the problem of how to evaluate density forecasts.[64] The basic idea is to check whether the actual realisations of the forecast variable are, on average, consistent with the forecast densities. To check this, Diebold, Gunther and Tay (1998) propose the estimation of the probability integral transform ($Z_t$) of the reported empirical densities. $Z_t$ is defined as the cumulative probability distribution function evaluated at the actual outcome, $X_t$, for the forecast period in question, i.e.

$$Z_t = \int_{-\infty}^{X_t} p(u_t)du \qquad \text{A3.1}$$

In other words, $Z_t$ gives the estimated probability of the variable being less than or equal to the actual outcome (observed only ex post) according to the predictive forecast density. Diebold, Gunther and Tay (1998) show how the properties of $Z_t$ depend on how closely the reported densities approximate the true underlying density driving the variable under study. In particular, if survey respondents accurately assess the true underlying probabilities and their evolution through time, Diebold, Gunther and Tay (1998) show that $Z_t$ will be a uniformly, independent and identically distributed random variable bounded between zero and unity, i.e. $Z_t$ uiid [0, 1].[65] Conversely, if forecasters have not accurately assessed the shape and location of the true density, the $Z_t$

64 For applications of the probability integral transform, see Diebold, Gunther and Tay (1998) who address the univariate case. Diebold, Hahn and Tay (1999) and Clements and Smith (2000) consider multivariate density forecasts, while Clements (2003) proposes an alternative approach based on derived event probability forecasts. Diebold, Tay and Wallis (1999) evaluate the density forecasts for inflation from the Federal Reserve Bank of Philadelphia's Survey of Professional Forecasters.

65 A uiid [0, 1] random variable is uniformly, identically and independently distributed over the interval [0, 1]. One important statistical feature implied by the independence criteria is that the variable will not be serially correlated.

series will display non-uniformities that highlight the "discrepancies" between the reported and true densities. For example, suppose that forecasters correctly assess the mean of future inflation. However, they mistakenly assume a Normal distribution while the true distribution actually has higher-than-normal probabilities of extreme outcomes (i.e. "fat tails"). Under such circumstances, respondents will not have assigned sufficient probability to outcomes in the tails of the distribution. Hence, there will tend to be many more realisations of inflation taking on extreme values than predicted by respondents' distributions. Consequently, reflecting the excessively low probabilities assigned by respondents to outcomes at the extreme of the distribution, there will also be a non-uniform clustering of $Z_t$ towards the extremes of the [0, 1] interval. In addition, reflecting the excessively high probability assigned by respondents to outcomes in the middle of the distribution, there will tend to be a non-uniform clustering of the $Z_t$ series towards the middle of the [0, 1] interval. In describing the distribution for the probability integral transforms for such non-optimal density forecasts, Diebold, Gunther and Tay (1998) refer to a "butterfly" shape.

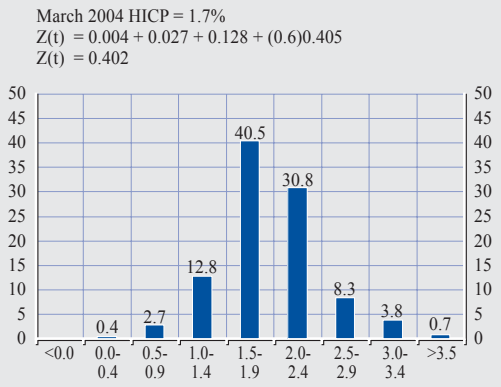
As regards the independence criterion, it is in line with the notion that forecasters are making use of available information sets, which include any past values of $Z_t$ that were available to respondents when the density forecasts were constructed. For one-step ahead forecast densities, this implies that the $Z_t$ series is independent, while for $h$-step ahead forecast densities, the $Z_t$ series should be, at most, $h$-$1$ dependent (i.e. there should be, at most, serial correlation of order $h$-$1$) (see Diebold, Gunther and Tay, 1998).

The SPF density forecasts can therefore be evaluated by assessing whether and how closely the empirical distribution of $Z_t$ corresponds to the uiid [0, 1] distribution that would arise under the assumption that the density forecasts were optimal.[66] One of the attractions of the above approach is that it requires no information on the model or other approach used by survey respondents to assess uncertainty. Rather, the analysis is based solely on the reported distributions relative to the realised inflation outcome. An additional advantage is that the above test of density forecasts is valid, even if the true densities are time varying. Notwithstanding these strengths, a robust assessment of density forecasts based on this approach requires relatively large samples which are often not available in macroeconomic contexts (see Pesaran and Skouras, 2002).

Aside from small sample issues, an additional practical difficulty also emerges in the current context of using the ECB SPF data to estimate $Z_t$. In particular, SPF respondents report discrete distributions which assign probabilities to various intervals that are specified in the survey questionnaire. Hence, given that the realised inflation outcome can sometimes fall within one of the surveyed intervals, it is necessary to approximate $Z_t$ in some way. For illustrative purposes, Chart 31 displays the approximation of $Z_t$ for the specific case of the probability distribution for the rate of inflation in March 2004, as surveyed in the second quarter of 2002 (i.e. the two-year ahead inflation forecast). As can be seen from the chart, respondents assigned most of the probability mass to outcomes within the 1.5%-1.9% range, while at the same time assigning a somewhat greater probability to outcomes above this range compared with the probability assigned to lower outcomes. $Z_t$, according to equation A3.1, provides the estimated probability – implied by the reported distribution – of inflation being below its realised value of 1.7%. This is calculated as the sum of the probabilities assigned to all ranges below 1.5% (i.e. 0.4%+2.7%+12.8%) plus some fraction of the probability assigned to the 1.5%-1.9% range (40.5%). In principle, the reported distribution does not provide precise information on how to estimate this fraction. In the empirical application below, this fraction is estimated by assuming that the total probability mass within

66 Peseran and Skouras (2002) point out that the uiid property represents a necessary, but not a sufficient, condition for the optimality of the underlying predictive distribution.

**Chart 31 Example: calculation of probability integral transform for the two-year ahead forecast of March 2004 inflation**




Source: ECB calculations.

the range is uniformly distributed across each of the five possible outcomes within that range. Hence, when inflation is rounded to one decimal point, of the five possible outcomes within the 1.5%-1.9% range, three are less than or equal to the actual inflation outcome in March 2004 (1.7%). This implies an estimate of 60% (or three-fifths) for the fraction of the probability that should be included in the calculation of $Z_t$.[67] As shown in the chart, $Z_t$ is then estimated to take on a value of 0.402 under these assumptions.

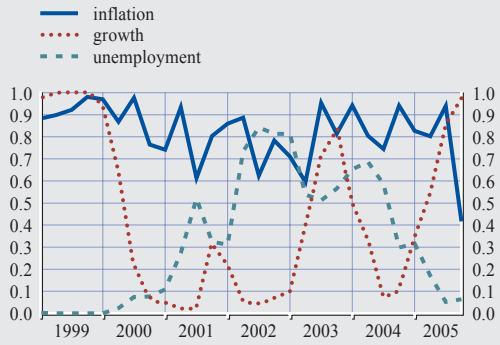
## EMPIRICAL APPLICATION TO THE ECB SPF

In this section, we report the estimated probability integral transforms for the ECB SPF aggregate density forecasts for inflation, GDP growth and the unemployment rate. The analysis is illustrated using the rolling one and two-year ahead density forecasts for each of the variables. For the one-year ahead forecast densities, the sample size ($n$) corresponds to only 28 observations for each variable, i.e. inflation, GDP growth and the unemployment rate, respectively. This is the sample size ($n$=28) considered by Diebold, Tay and Wallis (1999); however, in that study, the observations were not overlapping as in this case. For the 2-year ahead forecasts, the corresponding number of observations is reduced further to only 24 observations. The relatively small sample size highlights the clearly illustrative nature of the exercise. It is also important to underline here that, because we focus on year-on-year rates, the observations overlap. Thus, for example, the year-on-year rate in the second quarter of 2004, which is the sum of the rate of change in first and second quarters of 2004 and the third and fourth quarters of 2003, to a large extent overlaps with the year-on-year rate in the first quarter of 2004, which is the sum of the rate of change in the first quarter of 2004 and the last three quarters of 2003. Thus, a shock in any specific quarter may affect the year-on-year rate for a number of quarters.

Thus, as discussed above, there is likely to be some serial dependence in the estimated $Z_t$ series, reflecting the multi-step (i.e. one or two-year ahead) nature of these density forecasts. Moreover, it is not known how deviations from the independence criteria impact on the uniformity property and vice versa. Ideally, under such circumstances, one way of proceeding (see Diebold, Gunther and Tay, 1998) is to split the series into $h$ sub-series that should be iid (e.g. using quarterly one-year ahead forecasts, one could have a series of forecasts for the first quarter of each year, a series of forecasts for the second quarter, etc.). However, the relatively small sample size prohibits any meaningful analysis of such sub-series at present. For this reason, we proceed in

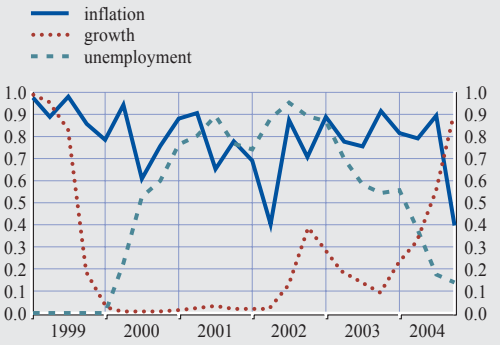
67 This approach to apportioning the probability within each bin or forecast range seemed the most consistent with the discrete nature of the probability distributions that respondents are asked to complete as part of the SPF questionnaire, and in the light of the fact that inflation is normally reported as a figure that is rounded to one decimal point. Of course, this should not be seen as ruling out different assumptions should the context require it. For example, in Section 3, the possibility of fitting continuous distributions (such as a Normal distribution) to the discrete SPF distributions in order to estimate the mean and the variance is considered.

**Chart 32 Probability integral transforms for one-year ahead inflation, GDP growth and unemployment rate density forecasts**




Source: ECB calculations.
Note: Horizontal axis refers to the survey date.

**Chart 33 Probability integral transforms for two-year ahead inflation, GDP growth and unemployment rate density forecasts**




Source: ECB calculations
Note: Horizontal axis refers to the survey date.

a more heuristic manner, examining the properties of $Z_t$ graphically.[68]

Chart 32 plots the estimated $Z_t$ series for inflation, real GDP growth and the euro area unemployment rate for the one-year ahead density forecasts for the surveys conducted since the first quarter of 1999. In the chart, large values of the $Z_t$ series correspond to unexpectedly high realised values for each of the three variables. From the chart, in the case of the one-year ahead inflation densities, there is some evidence to suggest that survey respondents have not accurately assessed the uncertainty surrounding future inflation outcomes. In particular, for all of the surveys conducted between the first quarter of 1999 and the fourth quarter of 2005, the estimated probability integral transform does not appear to be uniformly distributed within the [0, 1] interval. Rather, all of the observations, save for the last one, fall within the upper half of the [0, 1] interval. This considerable persistence is in line with the previous finding that respondents inaccurately assessed the centre or location of the true density (i.e. their inflation forecast tended to be biased downwards). In other words, SPF respondents failed to attribute sufficiently high probabilities to outcomes for inflation above the actual realised values. A similar picture emerges in the case of the two-year ahead inflation densities (reported in Chart 33), with only two observations from the surveys for the second quarter of 2002 and the fourth quarter of 2004 falling into the lower half of the [0, 1] interval. Using a much longer time series for US inflation (from 1970 to the late 1990s), it is interesting to note that Diebold, Tay and Wallis came to a similar conclusion; namely that positive (negative) inflation surprises occur more (less) often than the respondents to the Federal Reserve Bank of Philadelphia's SPF expect.

Turning to the assessment of the GDP growth and unemployment rate densities, the estimated probability integral transforms suggest a somewhat different picture to that emerging for inflation. In particular, in the case of both variables, even in this relatively small sample, observed values of $Z_t$ appear to be drawn from

68 We also assess the extent of serial correlation in the estimated probability integral transforms using Ljung-Box Q statistics (not reported). The results confirm the presence of significant serial correlation in line with what would be expected given the multi-step nature of the forecasts. In the case of the one-year ahead forecasts, the serial dependence is particularly evident for the probability integral transforms from the unemployment and GDP growth rate densities for all lags up to six (the maximum considered given the small sample). For inflation, the evidence of serial correlation is weaker, particularly for orders greater than two. In the case of the two-year ahead densities, the evidence of serial dependence is weaker for both GDP growth and inflation, while, for the unemployment rate densities, considerable serial correlation remains.

both the upper and lower halves of the [0, 1] interval, although at first glance they do not appear independent. In the first year of the survey (1999), respondents appeared to have underestimated the probabilities of high growth outcomes, but underestimated the probabilities of lower unemployment. However, this situation changed in surveys from early 2000 onwards. For example, for both the one and two-year ahead density forecasts for GDP growth, the estimated $Z_t$ clusters around zero for most of the period from end-2000 to the beginning of 2003, suggesting that respondents are likely to have underestimated the probabilities of lower GDP growth over this period. Similarly, in the case of the unemployment rate densities, the clustering and persistence of the probability integral transform in the upper half of the [0, 1] interval for surveys conducted between 2001 and 2004 may reflect a tendency among respondents to underestimate the probability of higher unemployment rates. This finding is of course consistent with the previous evidence that they may have underestimated the risks associated with the general economic slowdown.

## CONCLUDING REMARKS

This Annex has considered the problem of how to evaluate the density forecasts contained in the ECB SPF. Using the probability integral transform suggested by Diebold, Gunther and Tay (1998), the one and two-year ahead densities for inflation, GDP growth and the unemployment rate are assessed using the density forecasts available from the surveys conducted since the first quarter of 1999. The analysis is clearly restricted given the small time series sample that is available. Nonetheless, some evidence is found to suggest that respondents failed to accurately assess the risk of higher inflation outcomes over this period. In addition, SPF respondents may also have failed to accurately assess the risks of lower growth and higher unemployment associated with the economic downturn. Overall, however, it would seem that more definitive and robust conclusions must await the accumulation of further data. In the meantime, an interesting area for further investigation is the evaluation of the density forecasts of individual panel members in order to assess whether some members of the panel are better and more accurately able to assess macroeconomic risks than others.

## REFERENCES

Baghestani, H. M. and A. M. Kianian (1993), "On the Rationality of US Macroeconomic Forecasts: Evidence from a Panel of Professional Forecasters", in *Applied Economics*, Vol. 25, 1993, pp. 869-78.

Batchelor, R., (2001), "How useful are the forecasts of intergovernmental agencies? The IMF and the OECD versus Consensus", in *Applied Economics*, Vol. 33, 2001, pp. 225-235.

Blanchard, O., (1997), "The Medium Run", in *Brookings Papers on Economic Activity*, 2, 1997, pp. 89-158.

Bordes, C. and Clerc L., (2004), "Price stability and the ECB's monetary policy strategy", Note d'études et de recherche de la Banque de France, No. 109, March 2004.

Brown, B. W. and Maital S., (1981), "What do economists know? An empirical study of experts' expectations", *Econometrica*, Vol. 49, No. 2, March 1981, pp. 491-504.

Carroll, C., (2001), "The epidemiology of macroeconomic expectations", NBER Working Paper No. 8695, December 2001.

Castelnuovo, E., Nicoletti-Altimari, S. and Rodriguez Palenzuela, D., (2003), "Definition of price stability, range and point inflation targets: the anchoring of long-term inflation expectations", ECB Working Paper series, No. 273, September 2003.

Clements, M. P., (2003), "Evaluating the Survey of Professional Forecasters probability distributions of expected inflation based on derived event probability forecasts", mimeo, 2003.

Clements, M. P. and Smith, J., (2000), "Evaluating the forecast densities of linear and nonlinear models: application to output growth and unemployment", in *Journal of Forecasting*, Vol. 19, pp. 255-276.

Crawford, A. and Kasumovich (1996), M., "Does Inflation Uncertainty Vary with the Level of Inflation?", Working Papers 96-09, Bank of Canada, 1996.

Croushore, D., (1997), "The Livingstone survey: still useful after all these years", in *Federal Reserve Bank of Philadelphia Business Review*, March-April 1997, pp. 1-12.

Croushore, D., (1993), "Introducing the Survey of Professional Forecasters", in *Federal Reserve Bank of Philadelphia Business Review*, November-December 1993, pp. 1-13.

D'Amico, S. and Orphanides, A., (2006), "Uncertainty and Disagreement in Economic Forecasting", mimeo, August 2006.

Diebold, F. X., Gunther, T. A. and Tay, A. S., (1998), "Evaluating Density Forecasts with Applications to Financial Risk Management", in *International Economic Review*, Vol. 39, No. 4, November 1998, pp. 863-883.

Diebold, F. X., Hahn, J. and Tay, A. S., (1999), "Multivariate density forecast evaluation and calibration in financial risk management: high frequency returns on foreign exchange", in *Review of Economics and Statistics*, Vol. 81, 1999, pp. 661-73.

Diebold, F. X., Tay, A. S. and Wallis, K., (1999), "Evaluating density forecasts of inflation: the survey of Professional Forecasters", in *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive W. J. Granger*, Engle, R. F. and White, H., eds., Oxford University Press, 1999, pp. 76-90.

ECB (2005a), Box entitled "Trends in euro area potential output growth" in the July 2005 issue of the Monthly Bulletin.

ECB (2005b), Box entitled "A longer-term perspective on structural unemployment in the euro area" in the August 2005 issue of the Monthly Bulletin.

ECB (2004a), The monetary policy of the ECB, 2nd edition, January 2004.

ECB (2004b), Article entitled "Extracting information from financial asset prices" in the November 2004 issue of the Monthly Bulletin.

ECB (2002), Box entitled "Deriving long-term euro area inflation expectations from index-linked bonds issued by the French Treasury" in the February 2002 issue of the Monthly Bulletin.

ECB (2000), Article entitled "Developments in and structural features of the euro area labour markets" in the May 2000 issue of the Monthly Bulletin.

European Commission (2006), Report by the Economic Policy Committee and the European Commission on the impact of ageing populations on public spending, February 2006.

Figlewski, S. and Wachtel, P., (1983), "Rational Expectations, Informational Efficiency and Tests using Survey Data: a Reply", in *Review of Economics and Statistics*, Vol. 65, 1983, pp. 529-531.

Garcia, J. A., (1993), "An introduction to the ECB's survey of professional forecasters", ECB Occasional Paper No. 8, September 2003.

Giordani, P. and Söderlind, P., (2003), "Inflation forecast uncertainty", in *European Economic Review*, Vol. 47, No. 6, 2003, pp. 1037-1059.

Golob, J. E., (1994), "Does inflation uncertainty increases with inflation?", in *Federal Reserve Bank of Kansas City Economic Review*, Third Quarter, 1994, pp. 27-38.

Hafer, R. W. and Resler D. H., (1980), "The rationality of survey-based inflation forecasts", in *Federal Reserve Bank of Saint-Louis Business Review*, November 1980, pp. 3-11.

Harri, A. and Brorsen, B. W., (2002), "The overlapping data problem", Working Paper, Oklahoma State University, 2002.

Holden, K. and Peel, D. A., (1990), "On Testing for Unbiasedness and Efficiency of Forecasts", *The Manchester School of Economic & Social Studies,* Vol. 58, Issue 2, pp. 120-27.

Issing, O., (2003), (ed.), "Background studies for the ECB's evaluation of its monetary policy strategy", ECB publication, November 2003.

Keane, M. P. and Runkle, D. E., (1990), "Testing the rationality of price forecasts: new evidence from panel data", in *American Economic Review*, Vol. 80, No. 4, September 1990, pp. 714-735.

Keynes, J. M., (1936), "The general theory of employment, interest and money", Macmillan Cambridge University Press.

Kiley, Michael T., (2000), "Endogenous Price Stickiness and Business Cycle Persistence", in *Journal of Money, Credit, and Banking*, Vol. 32, Issue 1, February 2000, pp. 28-53.

Kontonikas, A., (2004), "Inflation and inflation uncertainty in the United Kingdom: evidence from GARCH modelling", in *Economic Modelling*, Elsevier, Vol. 21, Issue 3, 2004, pp. 525-543.

Kwan, S., (2005), "Inflation expectations: how the market speaks", in *Federal Reserve Bank of San Francisco Economic Letter*, 2005-25, 3 October.

Lahiri, K., and Liu F., (2006), "ARCH Models for Multi-period Forecast Uncertainty – A Reality Check Using a Panel of Density Forecasts", in *Econometric Analysis of Financial and Economic Time Series – Part A*, Terrell, D. and Fomby, T. B. (Eds.), Elsevier, JAI Press, 2006.

Lahiri, K., and Liu, F., (2004), "Determinants of multi-period forecast uncertainty using a panel of density forecasts", Econometric Society 2004 Australasian Meetings, No. 329, 2004.

Lahiri, K., Teigland, C. and Zaporowski, M., (1988), "Interest rates and the subjective probability distribution of inflation forecasts", in *Journal of Money, Credit and Banking*, Vol. 20, Issue 2, Ohio University Press, 1988, pp. 233-248.

Linden, S., (2003), "Assessment of GDP forecast uncertainty", European Commission Directorate General for Economics and Financial Affairs Economic Papers, No. 184, May 2003.

Muth, J. F., (1961), "Rational Expectations and the Theory of Price Movements", reprinted in The new classical macroeconomics. Volume 1. (1992): 3-23 (International Library of Critical Writings in Economics, Vol. 19. Aldershot, United Kingdom: Elgar).

Pesaran, M. H. and Skouras, S., (2002), "Decision-based methods for forecast evaluation", in A *Companion to Economic Forecasting*, Clements, M. P. and Hendry, D. F. (eds.), Blackwells, Oxford, 2002, pp. 268-283.

Sargent, T. J., (1993), "Bounded rationality in macroeconomics", Clarendon Press, Oxford, United Kingdom, 1993.

Sims, C., (2002), "The role of Models and Probabilities in the Monetary Policy Process", in *Brookings Papers on Economic Activity*, 2: 2002, pp. 1-62.

Stuart, A., and Ord, J. K., (1994), "Kendall's advanced theory of statistics", Vol. 1, Halsted Press, New York 1994.

Tay, A. and Wallis, K. F., (2000), "Density forecasting: a survey", Econometric Society World Congress 2000, Contributed Papers 0370, Econometric Society.

Taylor, J. B., (1999), "Monetary policy rules", NBER and University of Chicago Press (eds.), Chicago, 1999.

Thomas, L. B., (1999), "Survey measures of expected US inflation", in *Journal of Economic Perspectives*, Vol. 13, No. 4, Fall 1999, pp. 125-144.

Timmermann, A., (2006), "An evaluation of the World Economic Outlook Forecasts", IMF Working Paper WP/06/59, March 2006.

Urich, T. J. and P. Wachtel, (1984), "The effects of inflation and money supply announcements on interest rates", in *Journal of Finance,* Vol. 39, No. 4, September 1984, pp. 1177-88.

Wallis, K. F., (2004), "An assessment of Bank of England and National Institute inflation forecast uncertainties, in *National Institute Economic Review*, No. 189, 2004, pp. 64-71.

Wallis, K. F., (2001), "Chi-squared test of interval and density forecast and the Bank of England's fan Charts", ECB Working Paper No. 83, November 2001.

Zarnowitz, V. and Lambros, L. A., (1987), "Consensus and Uncertainty in Economic Prediction", in *Journal of Political Economy*, Vol. 95, No. 3, 1987, pp. 591-621.

## EUROPEAN CENTRAL BANK OCCASIONAL PAPER SERIES

1 "The impact of the euro on money and bond markets" by J. Santillán, M. Bayle and C. Thygesen, July 2000.

2 "The effective exchange rates of the euro" by L. Buldorini, S. Makrydakis and C. Thimann, February 2002.

3 "Estimating the trend of M3 income velocity underlying the reference value for monetary growth" by C. Brand, D. Gerdesmeier and B. Roffia, May 2002.

4 "Labour force developments in the euro area since the 1980s" by V. Genre and R. Gómez-Salvador, July 2002.

5 "The evolution of clearing and central counterparty services for exchange-traded derivatives in the United States and Europe: a comparison" by D. Russo, T. L. Hart and A. Schönenberger, September 2002.

6 "Banking integration in the euro area" by I. Cabral, F. Dierick and J. Vesala, December 2002.

7 "Economic relations with regions neighbouring the euro area in the 'Euro Time Zone'" by F. Mazzaferro, A. Mehl, M. Sturm, C. Thimann and A. Winkler, December 2002.

8 "An introduction to the ECB's survey of professional forecasters" by J. A. Garcia, September 2003.

9 "Fiscal adjustment in 1991-2002: stylised facts and policy implications" by M. G. Briotti, February 2004.

10 "The acceding countries' strategies towards ERM II and the adoption of the euro: an analytical review" by a staff team led by P. Backé and C. Thimann and including O. Arratibel, O. Calvo-Gonzalez, A. Mehl and C. Nerlich, February 2004.

11 "Official dollarisation/euroisation: motives, features and policy implications of current cases" by A. Winkler, F. Mazzaferro, C. Nerlich and C. Thimann, February 2004.

12 "Understanding the impact of the external dimension on the euro area: trade, capital flows and other international macroeconomic linkages" by R. Anderton, F. di Mauro and F. Moneta, March 2004.

13 "Fair value accounting and financial stability" by a staff team led by A. Enria and including L. Cappiello, F. Dierick, S. Grittini, A. Maddaloni, P. Molitor, F. Pires and P. Poloni, April 2004.

14 "Measuring financial integration in the euro area" by L. Baele, A. Ferrando, P. Hördahl, E. Krylova, C. Monnet, April 2004.

15 “Quality adjustment of European price statistics and the role for hedonics” by H. Ahnert and G. Kenny, May 2004.

16 “Market dynamics associated with credit ratings: a literature review” by F. Gonzalez, F. Haas, R. Johannes, M. Persson, L. Toledo, R. Violi, M. Wieland and C. Zins, June 2004.

17 “Corporate ‘excesses’ and financial market dynamics” by A. Maddaloni and D. Pain, July 2004.

18 “The international role of the euro: evidence from bonds issued by non-euro area residents” by A. Geis, A. Mehl and S. Wredenborg, July 2004.

19 “Sectoral specialisation in the EU: a macroeconomic perspective” by MPC task force of the ESCB, July 2004.

20 “The supervision of mixed financial services groups in Europe” by F. Dierick, August 2004.

21 “Governance of securities clearing and settlement systems” by D. Russo, T. Hart, M. C. Malaguti and C. Papathanassiou, October 2004.

22 “Assessing potential output growth in the euro area: a growth accounting perspective” by A. Musso and T. Westermann, January 2005.

23 “The bank lending survey for the euro area” by J. Berg, A. van Rixtel, A. Ferrando, G. de Bondt and S. Scopel, February 2005.

24 “Wage diversity in the euro area: an overview of labour cost differentials across industries” by V. Genre, D. Momferatou and G. Mourre, February 2005.

25 “Government debt management in the euro area: recent theoretical developments and changes in practices” by G. Wolswijk and J. de Haan, March 2005.

26 “The analysis of banking sector health using macro-prudential indicators” by L. Mörttinen, P. Poloni, P. Sandars and J. Vesala, March 2005.

27 “The EU budget – how much scope for institutional reform?” by H. Enderlein, J. Lindner, O. Calvo-Gonzalez, R. Ritter, April 2005.

28 “Reforms in selected EU network industries” by R. Martin, M. Roma, I. Vansteenkiste, April 2005.

29 “Wealth and asset price effects on economic activity”, by F. Altissimo, E. Georgiou, T. Sastre, M. T. Valderrama, G. Sterne, M. Stocker, M. Weth, K. Whelan, A. Willman, June 2005.

30 “Competitiveness and the export performance of the euro area”, by a Task Force of the Monetary Policy Committee of the European System of Central Banks, June 2005.

31 “Regional monetary integration in the member states of the Gulf Cooperation Council (GCC)” by M. Sturm and N. Siegfried, June 2005.

32 "Managing financial crises in emerging market economies: experience with the involvement of private sector creditors" by an International Relations Committee task force, July 2005.

33 "Integration of securities market infrastructures in the euro area" by H. Schmiedel, A. Schönenberger, July 2005.

34 "Hedge funds and their implications for financial stability" by T. Garbaravicius and F. Dierick, August 2005.

35 "The institutional framework for financial market policy in the USA seen from an EU perspective" by R. Petschnigg, September 2005.

36 "Economic and monetary integration of the new Member States: helping to chart the route" by J. Angeloni, M. Flad and F. P. Mongelli, September 2005.

37 "Financing conditions in the euro area" by L. Bê Duc, G. de Bondt, A. Calza, D. Marqués Ibáñez, A. van Rixtel and S. Scopel, September 2005.

38 "Economic reactions to public finance consolidation: a survey of the literature" by M. G. Briotti, October 2005.

39 "Labour productivity in the Nordic EU countries: a comparative overview and explanatory factors – 1998-2004" by A. Annenkov and C. Madaschi, October 2005.

40 "What does European institutional integration tell us about trade integration?" by F. P. Mongelli, E. Dorrucci and I. Agur, December 2005.

41 "Trends and patterns in working time across euro area countries 1970-2004: causes and consequences" by N. Leiner-Killinger, C. Madaschi and M. Ward-Warmedinger, December 2005.

42 "The New Basel Capital Framework and its implementation in the European Union" by F. Dierick, F. Pires, M. Scheicher and K. G. Spitzer, December 2005.

43 "The accumulation of foreign reserves" by an International Relations Committee Task Force, February 2006.

44 "Competition, productivity and prices in the euro area services sector" by a Task Force of the Monetary Policy Committee of the European System of Central banks, April 2006.

45 "Output growth differentials across the euro area countries: Some stylised facts" by N. Benalal, J. L. Diaz del Hoyo, B. Pierluigi and N. Vidalis, May 2006.

46 "Inflation persistence and price-setting behaviour in the euro area – a summary of the IPN evidence", by F. Altissimo, M. Ehrmann and F. Smets, June 2006.

47 "The reform and implementation of the stability and growth pact" by R. Morris, H. Ongena and L. Schuknecht, June 2006.

48 “Macroeconomic and financial stability challenges for acceding and candidate countries” by the International Relations Committee Task Force on Enlargement, July 2006.

49 “Credit risk mitigation in central bank operations and its effects on financial markets: the case of the Eurosystem” by U. Bindseil and F. Papadia, August 2006.

50 “Implications for liquidity from innovation and transparency in the European corporate bond market” by M. Laganá, M. Peřina, I. von Köppen-Mertes and A. Persaud, August 2006.

51 “Macroeconomic implications of demographic developments in the euro area” by A. Maddaloni, A. Musso, P. Rother, M. Ward-Warmedinger and T. Westermann, August 2006.

52 “Cross-border labour mobility within an enlarged EU” by F. F. Heinz and M. Ward-Warmedinger, October 2006.

53 “Labour productivity developments in the euro area” by R. Gomez-Salvador, A. Musso, M. Stocker and J. Turunen, October 2006.

54 “Quantitative quality indicators for statistics – an application to euro area balance of payment statistics” by V. Damia and C. Picón Aguilar, November 2006.

55 “Globalisation and euro area trade: Interactions and challenges” by U. Baumann and F. di Mauro, February 2007.

56 “Assessing fiscal soundness: Theory and practice” by N. Giammarioli, C. Nickel, P. Rother, J.-P. Vidal, March 2007.

57 “Understanding price developments and consumer price indices in south-eastern Europe” by S. Herrmann and E. K. Polgar, March 2007.

58 “Long-term growth prospects for the Russian economy” by R. Beck, A. Kamps and E. Mileva, March 2007.

59 “The ECB survey of professional forecasters (SPF) – A review after eight years’ experience” by C. Bowles, R. Friz, V. Genre, G. Kenny, A. Meyler andT. Rautanen, March 2007.

ISSN 1607148-4