INTERNATIONAL ORGANISATIONS' *VS.* PRIVATE ANALYSTS' GROWTH FORECASTS: AN EVALUATION*

*Ildeberta Abreu***

ABSTRACT

This article evaluates the performance of economic growth forecasts disclosed by three international organisations – the IMF, the European Commission and the OECD – and compares it with that of the mean forecasts of two surveys of private analysts – the Consensus Economics and The Economist. The aim is to help forecast users in answering the question of how much (little) confidence they should place in the alternative forecasts that are available at each moment. The evaluation covers projections for nine advanced economies over the period 1991-2009. Several evaluation criteria are used: the quantitative and the directional accuracy of forecasts and, also, the ability to predict economic recessions. The results suggest that the forecasting performance of the international organisations is broadly similar to that of the surveys of private analysts. By and large, current-year forecasts present desirable features and clearly outperform year-ahead forecasts for which evidence is more mixed.

1. Introduction

Considerable effort and resources are devoted to forecasting major economic variables and the publication of forecasts usually attracts great interest of economists, policymakers and the general public. Although some of the disappointment that arises from time to time with macroeconomic forecasting might be justified, part of it reflects a failure to inform forecast users of how much (little) confidence to place in forecasts. An empirical evaluation of the past accuracy of the various forecasters and of their relative performance might help the user to make an informed use of the many different predictions available.

This article will evaluate the forecasting record of three leading international organisations – the International Monetary Fund (IMF), the European Commission (EC) and the Organisation for Economic Co-operation and Development (OECD) – and compare it with that of two surveys of private analysts – the Consensus Economics and The Economist. The forecasts published twice a year by the three international organisations receive a great deal of media attention and are usually perceived to benefit from the large amount of intellectual/physical resources devoted to their production. However, many private sector analysts (including banks, corporations, consultants, etc.) also produce forecasts making use of their knowledge about the countries where they are based. These private analysts' forecasts are published on a monthly basis and have been gaining increased visibility. In this analysis, and unlike most

* The author gratefully acknowledges Carlos R. Marques for the many valuable discussions and suggestions. The author also thanks Rita Duarte and João Sousa for their useful comments. The opinions expressed in the article are those of the author and do not necessarily coincide with those of the Banco de Portugal or the Eurosystem. Any errors and omissions are the sole responsibility of the author.

** Banco de Portugal, Economics and Research Department.

previous work on forecast evaluation, we want to place ourselves in the position of a forecast user that needs to know how much confidence to place on each of these various forecasts that are available at a specific point in time. Besides following a slightly different empirical approach for choosing the timing of comparison of the various forecasts, this analysis aims to contribute to the existing literature by assessing a less known survey of private forecasters (The Economist) and by extending the assessment to the most recent vintages of projections up to the latest recession.

The evaluation covers real Gross Domestic Product (GDP) growth forecasts, for the period 1991-2009, for nine main advanced economies.[1] Several evaluation criteria will be used. We will assess the accuracy of forecasts in terms of magnitude (quantitative accuracy) and test their unbiasedness and efficiency. We will also examine accuracy in terms of direction of change (directional accuracy) and briefly assess the ability of forecasters to predict economic recessions. The performance of forecasters will be judged against different benchmarks: firstly, against a "naive" benchmark which establishes a minimum level of accuracy that a forecast should have and, secondly, the accuracy of international organisations' forecasts will be compared to that of the alternative private analysts' forecasts. As much as possible, the statistical significance of these differences in accuracy will be tested.

The article is structured as follows. Section 2 describes in detail the data set and conventions used. Section 3 evaluates the quantitative accuracy of forecasts. The weak form efficiency of forecasts is studied in the following section. Section 5 examines two additional dimensions of accuracy: the directional accuracy and the ability to predict economic recessions. The last section summarises the results and briefly compares them with the findings of previous in-house evaluations of international organisations' forecasts.

2. Data set used

The study examines two groups of forecasts: the ones published by the IMF, the EC and the OECD and the mean forecasts of the panels of private analysts surveyed by the Consensus Economics and The Economist.[2] We make use of the fact that international organisations publish projections two times per year (generally, in Spring and in Autumn) for both the current-year and the year-ahead.[3] This means that we use four sets of forecasts which correspond to four different forecasting horizons. For a target year t, we will be looking at the Spring and Autumn next-year forecasts (reported in year $t-1$) and the Spring and Autumn current-year forecasts (reported in year t). For example, the IMF reported four forecasts for the 2000 German GDP growth: the Spring and Autumn 1999 next-year forecasts and the Spring and Autumn 2000 current-year forecasts. These forecasting horizons can be thought of as corresponding roughly to seven, five, three and one quarter-ahead, respectively.

To investigate the relative performance of international organisations and private analysts it is necessary to decide on the timing of the comparison given that the surveys of private analysts are available on a monthly basis. A valid argument would be to choose a reference month for which the information set underlying the private analysts' forecasts is similar to the one underlying each international organisa-

1 This article draws heavily on the work carried out in Abreu (2011), which also covers the performance of inflation forecasts. For additional details see *Working Paper* 20.

2 IMF, "World Economic Outlook"; EC, "European Economic Forecast"; OECD, "OECD Economic Outlook"; Consensus Economics, "Consensus Forecasts" and The Economist, "The Economist pool of forecasters".

3 We will not consider any interim assessments published by these organisations and neither the two-year-ahead forecasts that are published in Autumn by the EC and the OECD. For an evaluation of OECD's two-year-ahead growth forecasts see Vuchelen and Gutierrez (2005).

tion's forecasts. Most previous work on forecast evaluation tries to follow this approach but typically end-up using rough approximations. Moreover, according to tentative evidence on the sensitivity of the relative performance of international organisations and private forecasters to changes in the dating, such as the one presented in Timmermann (2007) and Lenain (2001), the timing of the comparison presumably matters.

We decided to follow a slightly different empirical strategy in this work. The idea is to place ourselves in the position of a user that has a new forecast just released by an international organisation and also the more recent forecasts released by private institutions and needs to have an informed judgement about their relative reliability. To do this, we first collected for each international organisation the public disclosure date of every forecasting exercise. Then, we selected for each private institution the forecast disclosed to the public at a closer date (before or no more than a couple of days after that of the international organisation). This means that the reference months used for the Consensus and for The Economist vary according to which international organisation they are being compared to and also differ somewhat over the sample period.[4]

The study focus on real GDP annual growth forecasts for nine advanced economies: the six major euro area countries (Germany, France, Italy, Spain, Netherlands and Belgium),[5] the United Kingdom, the United States and Japan. The set of countries was chosen both on account of their importance in the world economy and of data availability across the institutions and the period under analysis.[6] Note that the definitions of the data collected can differ across institutions and over time (*e.g.* the working-day adjustment of GDP data and the German reunification). As much as possible, given data availability, these differences are properly taken into account so that they do not affect the size of the forecast error. The observation period covers around two decades, from 1991 to 2009.[7] However, it is important to be aware that the relatively small sample size (19 observations at most for each forecasting horizon) may limit the robustness of the inference that can be made and the number of cyclical fluctuations to be studied.

Given that GDP data are subject to revisions, a choice has to be made concerning the outcome data to be used in the forecast evaluation. Though no single choice is optimal, we decided to take the conventional view that forecasters should be judged by their ability to predict the early releases of data rather than the later revisions, which often incorporate methodological changes and information that was not available to them at the time of forecasting.[8] Hence, for each institution we use as outcome value for year t the first-available data reported in their Spring forecast exercise of the following year $(t+1)$.[9] This choice has the additional advantage of allowing us to take into account the differences in definitions among institutions.

---

4 Roughly speaking, the reference months used were mostly April and September for comparison with the IMF, April/May and October/November for comparison with the EC and May/June and November/December for comparison with the OECD.

5 Which represent over 85 per cent of euro area GDP.

6 In particular, The Economist's survey does not provide forecasts for smaller euro area countries, including Portugal.

7 The forecast exercises analysed go from Autumn 1991 till Autumn 2009. In the case of the IMF's forecasts for Spain, Netherlands and Belgium the sample is slightly smaller given the lack of a couple of observations at the beginning of the period.

8 See McNees (1992) and Zarnowitz and Braun (1993) for a discussion on this issue.

9 In the case of private analysts, which no longer report year t data in their first forecast exercise of the following year, the outcome of one of the international organisations is used.

In this study, the forecast error (e) is defined as the difference between the outcome/actual value (y) and the forecasted value $(\hat{y})$. For each target year t, we analyse four different forecast errors corresponding to four different forecasting horizons (h). According to this notation, the forecast error can be generally written as:

$$e_{t,h} = y_t - \hat{y}_{t,h} \tag{1}$$

and the following designation will be used for the four different forecast errors:

$$e_{t,Spring_{t-1}} = y_t - \hat{y}_{t,Spring_{t-1}} \quad \text{Spring next-year forecast error}$$
$$e_{t,Autumn_{t-1}} = y_t - \hat{y}_{t,Autumn_{t-1}} \quad \text{Autumn next-year forecast error}$$
$$e_{t,Spring_t} = y_t - \hat{y}_{t,Spring_t} \quad \text{Spring current-year forecast error}$$
$$e_{t,Autumn_t} = y_t - \hat{y}_{t,Autumn_t} \quad \text{Autumn current-year forecast error}$$

3. Quantitative accuracy of forecasts

To evaluate the quantitative accuracy of forecasts we examine the forecast errors and compute a set of conventional summary measures. The aim is to characterize in a simple way the distribution of errors. The first measure is the mean error (ME), *i.e.* the arithmetic average of forecast errors over the available observations (n), for each horizon (h). Even though positive and negative errors might offset each other, the ME gives an indication of a possible bias in the forecasts, with a negative sign indicating an over-prediction on average of the actual value.

$$ME_h = \frac{1}{n}\sum_{t=1}^{n} e_{t,h} \tag{2}$$

The second is the standard deviation of errors (SD), which can give an indication about the uncertainty at each forecasting horizon.

$$SD_h = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}\left(e_{t,h} - ME_h\right)^2} \tag{3}$$

The third one is the root mean squared error $(RMSE)$, which is the square root of the sample average of squared forecast errors (*i.e.* the square root of the mean squared error (MSE)). The $RMSE$ disregards the sign of errors (puts equal weight on over- and under-predictions) and implicitly assumes that the seriousness of any error increases sharply with square the size of the error. Therefore, it penalises forecasters who make large errors.[10]

10 The $RMSE$ is consistent with a symmetric quadratic loss function of forecasters. This assumption will be discussed in Section 4.

$$RMSE_h = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_{t,h}^2} \qquad (4)$$

These measures have been subject to some criticisms (see, for example, Fildes and Stekler (2002)). The $RMSE$ can be particularly affected by outliers which are common in economic data sets. Also, neither the ME nor the $RMSE$ are scale independent. As done in Koutsogeorgopoulou (2000), we will adjust the $RMSE$ by the standard deviation of outcomes to compare performance across countries, in order to take into account the variability of the series being forecasted.

In addition, to evaluate the performance of a forecaster, these descriptive statistics are compared to similar statistics obtained from alternative forecasts available to the user. The first alternative is a "naive" benchmark that serves to establish a minimum level of accuracy that a forecast should have. A frequent procedure is to use a no-change naive model. In this work we use instead a same-change naive model, which extrapolates a GDP growth rate similar to the one observed in the last period. As argued by McNees (1992), this is a more stringent and sensible basis of comparison for variables that tend to grow over time. To be fair to forecasters, we use for each forecasting horizon the last rate of change known at the time of forecasting. This is similar to assume that the variable to be forecasted follows a random walk.[11] To formalise the comparison, we compute a version of Theil's inequality coefficient (U), defined as the ratio of the MSE of the forecaster being evaluated to the MSE of the naive forecast $\left(\hat{y}_{t,h}^{N}\right)$.[12] If the Theil's U is less than one the forecaster being evaluated beats the naive model. This measure, unlike others, is not affected by the units of measurement of data.

$$U_h = \frac{\frac{1}{n}\sum_{t=1}^{n}\left(y_t - \hat{y}_{t,h}\right)^2}{\frac{1}{n}\sum_{t=1}^{n}\left(y_t - \hat{y}_{t,h}^{N}\right)^2} \qquad (5)$$

The second alternative is the benchmarking of other experts' forecasts. In this study, the focus is on the comparison of the performance of each international organisation with that of the two private institutions. The comparison is based on the ratio of their respective $RMSE$.[13] A ratio higher than one indicates a lower accuracy of the international organisation relative to the private institution.

Irrespectively of the benchmark used to evaluate the performance of a forecaster, it is necessary to test whether a forecaster's errors are significantly different from those of the benchmark, *i.e.* the difference should be tested for statistical significance. For this purpose, we run the test for equal forecast accuracy proposed by Diebold and Mariano (1995). To implement the test we estimate the following equation:[14]

---

11 In practice this means that: in Spring and Autumn $t-1$, the naive forecast for growth in year t corresponds to the actual growth rate in year $t-2$; in Spring and Autumn t, the naive forecast corresponds to the actual growth rate in year $t-1$.

12 In the case of a no-change naive model, the Theil's U corresponds to the ratio of the MSE of the forecaster to the mean of squared outcomes, as originally proposed by Theil (1971).

13 Note that this ratio is equivalent to the square root of a corresponding Theil's U coefficient.

14 By ordinary least squares, using the Newey-West covariance estimator that is consistent in the presence of both heteroskedasticity and autocorrelation.

$$d_{t,h} = \alpha + \varepsilon_{t,h} \quad \text{where} \quad d_{t,h} = e_{t,h}^2 - e_{t,h}^{*\,2} \tag{6}$$

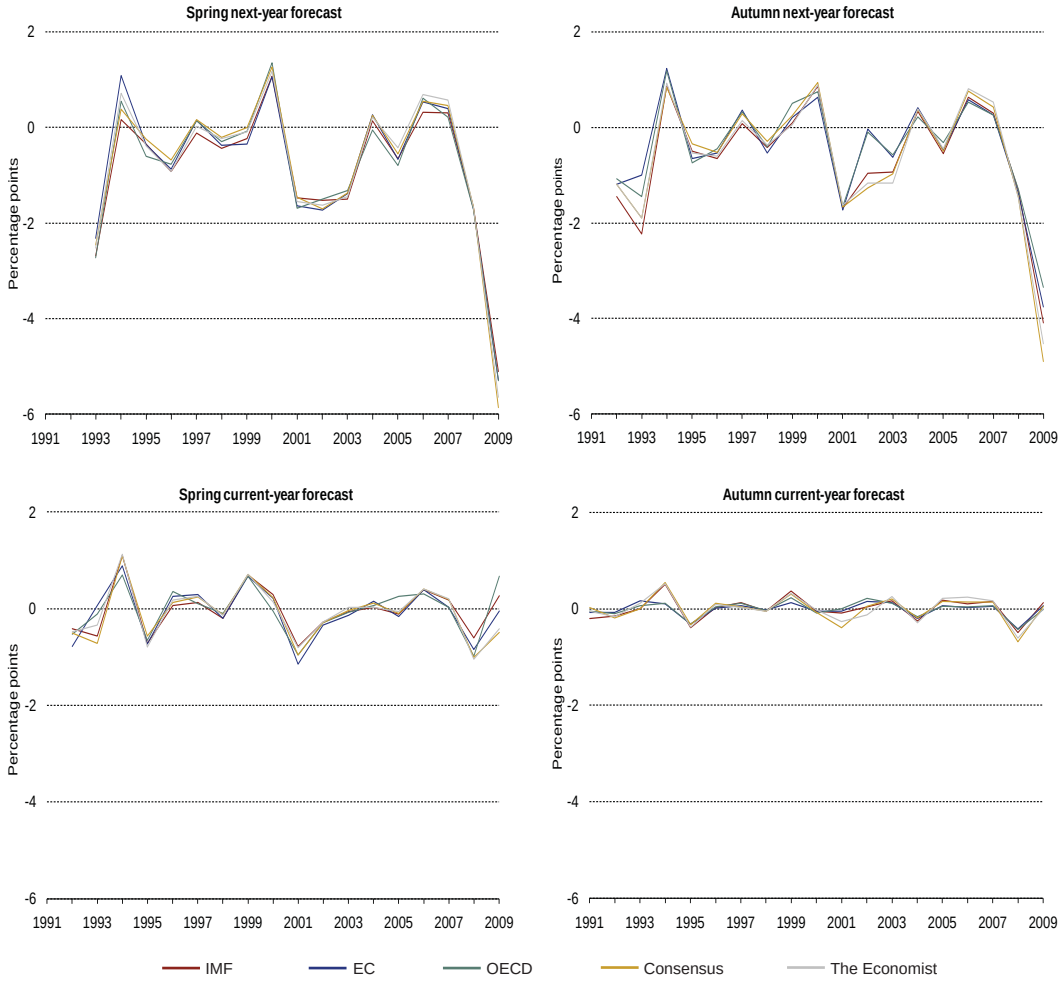
being $e_{t,h}$ the forecast errors of the forecaster being evaluated and $e_{t,h}^{*}$ the forecast errors of the benchmark (either the naive forecast or another forecaster). The null hypothesis of equal forecast accuracy $\left(H_0 : \alpha = 0\right)$ is tested using the small sample modifications proposed by Harvey *et al.* (1997).

3.1. A general look at forecast errors

Chart 1 provides a general picture of GDP growth forecast errors over time for each projection horizon.[15] For the sake of simplicity, data refer to the average of the nine countries under analysis but similar assertions hold at the individual country level.[16] It is clear that, for all institutions, errors are more significant for next-year forecasts and much closer to zero for current-year forecasts, especially

Chart 1

GDP GROWTH - FORECAST ERRORS FOR THE AVERAGE OF 9 COUNTRIES



Sources: Consensus Economics, EC, IMF, OECD, The Economist and author's calculations.

15 When presenting isolated data for the Consensus and The Economist they always correspond to the data set specifically used for comparison with the IMF's forecasts. Nothing in substance would change if the data sets used for comparison with the EC or the OECD were chosen instead.

16 See Abreu (2011) for information at the country level.

for the shorter projection horizon (Autumn current-year). Indeed, the profiles of next-year forecasts are generally flatter than the outcome while current-year forecasts tend to follow more closely the volatility of GDP growth. Forecast errors are quite similar across institutions as their forecasts tend to move closely together, particularly for current-year horizons.[17] The correlation coefficient of the various institutions' current-year forecasts for GDP growth is close to one.

Year-ahead forecast errors are predominantly below zero (overestimation) for most countries and are especially pronounced at the beginning and end of the sample period, when most countries were experiencing economic recessions. There is a tendency of the various forecasters to overestimate growth when activity is slowing down and, for most countries, this was stronger than the underestimation during upswings of economic activity.[18] Regarding current-year forecast errors, as mentioned before, they fluctuate around zero and do not seem to present a clear bias over the sample period.

Table 1 reports some summary statistics of projection errors. For the various countries and institutions, it is clear that accuracy improves as more relevant information becomes available to the forecaster. Both the ME and the $RMSE$ tend to be smaller as the horizon shortens. This is also true for the standard deviation of forecast errors and the reduction in uncertainty seems to be especially large as we move from next-year to current-year horizons. Regarding year-ahead horizons, the ME for the group of nine countries analysed is negative for all institutions. In fact, GDP growth was overestimated more than 50 per cent of the time by all forecasters. The mean error stands at around -0.8 p.p. of GDP growth for forecasts made in Spring $t-1$ and around -0.5 p.p. for forecasts made in Autumn $t-1$.[19] Given that actual GDP growth averaged 1.6 per cent a year over this period, the accuracy of year-ahead forecasts is not particularly impressive. The countries with larger mean errors are the three major euro area countries and Japan.[20] Let's just mention that the large negative mean error in the case of Japan is associated with a high standard deviation. Regarding current-year horizons, forecasts seem to be generally unbiased. For the group of countries studied, the mean forecast error is very small and in the case of Autumn current-year forecasts is basically zero.

Looking at the $RMSE$ adjusted by the standard deviation of GDP growth outcomes, to take into account the fact that countries with higher GDP volatility might be harder to predict, the forecasting performance becomes somewhat more similar across the various countries.

Finally, it should be mentioned that the correlation of projection errors across countries is higher for year-ahead horizons but especially among euro area countries and, though less so, among these and the United Kingdom. The United States' and Japan's forecast errors are weakly correlated with each other and with those of other countries. Therefore, it can be said that error correlation appears to be substantial only for longer horizons and for economies with more synchronised business cycles, such as the euro area countries.

---

17 As mentioned before, we decided to use for each institution its own outcome value (as reported in its Spring forecast exercise of the following year) but the outcomes for each country turn out to be quite similar across institutions.

18 This looks consistent with existing evidence of a considerable sluggishness in revisions of growth forecasts, as documented for example in Loungani *et al.* (2011).

19 If we exclude the 2009 recession, the mean error would still be negative but slightly less: around -0.5 p.p. for forecasts made in Spring $t-1$ and around -0.3 p.p. for forecasts made in Autumn $t-1$.

20 The statistical significance of the mean errors will be tested in Section 4.

Table 1 (to be continued)

DESCRIPTIVE STATISTICS OF GDP GROWTH FORECAST ERRORS (1991-2009)

	Memo:	Spring next-year forecast					Autumn next-year forecast					Spring current-year forecast					Autumn current-year forecast				
	Actual GDP growth	IMF	EC	OECD	Consen-sus	The Econo-mist	IMF	EC	OECD	Consen-sus	The Econo-mist	IMF	EC	OECD	Consen-sus	The Econo-mist	IMF	EC	OECD	Consen-sus	The Econo-mist
ME																					
Germany	*1.3*	-1.2	-1.0	-1.1	-1.0	-1.1	-0.9	-0.6	-0.6	-0.8	-0.9	0.1	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0
France	*1.5*	-1.0	-0.9	-0.9	-0.9	-0.9	-0.7	-0.5	-0.5	-0.7	-0.7	-0.1	-0.1	-0.1	-0.2	-0.1	-0.1	0.0	0.0	-0.1	-0.1
Italy	*0.8*	-1.4	-1.4	-1.3	-1.3	-1.3	-1.1	-1.0	-0.8	-1.1	-1.1	-0.4	-0.5	-0.3	-0.5	-0.4	-0.1	-0.1	-0.1	-0.2	-0.1
Spain	*2.3*	-0.6	-0.6	-0.6	-0.4	-0.3	-0.2	-0.4	-0.3	-0.2	-0.2	0.1	0.0	0.1	0.1	0.1	0.1	0.0	0.0	0.1	0.1
Netherlands	*1.7*	-0.7	-0.7	-0.8	-0.7	-0.6	-0.5	-0.3	-0.3	-0.5	-0.4	0.0	0.0	0.1	0.0	0.0	0.1	0.1	0.0	0.2	0.2
Belgium	*1.6*	-0.6	-0.8	-0.8	-0.6	-0.5	-0.3	-0.5	-0.5	-0.4	-0.4	0.1	-0.1	0.0	-0.1	0.0	0.1	0.0	0.1	0.1	0.1
United Kingdom	*1.6*	-0.6	-0.6	-0.6	-0.5	-0.6	-0.5	-0.4	-0.3	-0.5	-0.5	-0.1	-0.1	0.0	-0.1	-0.1	0.0	0.0	0.0	0.0	0.0
United States	*2.6*	0.0	0.2	0.0	-0.1	0.0	0.2	0.4	0.3	0.0	0.1	0.3	0.3	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0
Japan	*0.9*	-1.4	-1.2	-1.3	-1.1	-1.3	-1.2	-0.9	-0.7	-0.9	-1.1	-0.1	-0.3	-0.2	0.0	-0.1	-0.1	-0.1	-0.1	-0.2	-0.2
Average of 9 countries	*1.6*	-0.8	-0.8	-0.8	-0.7	-0.7	-0.6	-0.5	-0.4	-0.6	-0.6	0.0	-0.1	0.0	-0.1	-0.1	0.0	0.0	0.0	0.0	0.0
SD																					
Average of 9 countries	*1.9*	1.8	1.8	1.8	1.9	1.9	1.5	1.4	1.4	1.6	1.6	0.8	0.8	0.8	0.8	0.8	0.4	0.3	0.3	0.5	0.5
Percentage of negative errors																					
Average of 9 countries		63.3	58.8	60.8	57.8	57.1	57.1	54.3	54.3	55.1	57.1	47.2	47.5	46.3	54.7	51.6	38.8	40.9	38.0	46.7	43.0
RMSE																					
Germany		2.3	2.3	2.2	2.3	2.3	1.9	1.6	1.5	2.0	2.0	0.8	0.8	0.8	0.7	0.8	0.3	0.2	0.2	0.3	0.4
France		1.5	1.6	1.7	1.7	1.7	1.3	1.1	1.1	1.5	1.4	0.6	0.6	0.6	0.7	0.7	0.3	0.3	0.3	0.4	0.4
Italy		2.1	2.1	2.1	2.2	2.2	1.8	1.7	1.4	1.9	1.9	0.9	0.8	0.7	0.9	0.8	0.5	0.3	0.3	0.5	0.4
Spain		1.6	1.8	1.7	1.7	1.7	1.1	1.2	1.1	1.3	1.2	0.5	0.6	0.5	0.6	0.6	0.3	0.1	0.2	0.2	0.3
Netherlands		2.0	2.0	2.0	2.1	2.1	1.7	1.5	1.4	1.8	1.8	0.9	0.9	0.9	0.9	0.9	0.5	0.4	0.3	0.5	0.6
Belgium		1.6	1.8	1.8	1.7	1.7	1.3	1.3	1.3	1.6	1.5	0.8	0.7	0.7	0.9	0.8	0.4	0.3	0.3	0.6	0.5
United Kingdom		1.8	1.8	1.7	1.8	1.8	1.5	1.3	1.3	1.6	1.5	0.7	0.7	0.6	0.8	0.8	0.3	0.2	0.3	0.3	0.3
United States		1.5	1.5	1.6	1.7	1.6	1.3	1.3	1.3	1.5	1.3	0.6	0.7	0.5	0.6	0.5	0.4	0.4	0.4	0.4	0.4
Japan		2.8	2.5	2.7	2.7	2.7	2.3	2.1	2.1	2.3	2.4	1.4	1.4	1.3	1.3	1.3	0.7	0.6	0.6	0.8	0.8
Average of 9 countries		1.9	2.0	2.0	2.0	2.0	1.6	1.5	1.4	1.7	1.7	0.8	0.8	0.8	0.8	0.8	0.4	0.3	0.3	0.5	0.5

Table 1 (continued)

DESCRIPTIVE STATISTICS OF GDP GROWTH FORECAST ERRORS (1991-2009)

	Memo:	Spring next-year forecast					Autumn next-year forecast					Spring current-year forecast					Autumn current-year forecast				
	Actual GDP growth	IMF	EC	OECD	Consen-sus	The Econo-mist	IMF	EC	OECD	Consen-sus	The Econo-mist	IMF	EC	OECD	Consen-sus	The Econo-mist	IMF	EC	OECD	Consen-sus	The Econo-mist
$RMSE$ / SD **of actual GDP growth**																					
Germany	1.1	1.2	1.1	1.2	1.2	1.0	0.8	0.8	1.0	1.0	0.4	0.4	0.4	0.4	0.4	0.2	0.1	0.1	0.2	0.2
France	1.1	1.1	1.1	1.1	1.1	1.0	0.8	0.7	1.0	1.0	0.4	0.5	0.4	0.5	0.5	0.2	0.2	0.2	0.3	0.3
Italy	1.1	1.2	1.1	1.2	1.2	1.0	0.9	0.8	1.0	1.0	0.5	0.4	0.4	0.5	0.5	0.3	0.2	0.2	0.3	0.2
Spain	0.8	0.9	0.9	0.9	0.9	0.6	0.6	0.6	0.7	0.7	0.3	0.3	0.3	0.3	0.3	0.1	0.1	0.1	0.1	0.1
Netherlands	1.0	1.0	1.0	1.0	1.0	0.9	0.8	0.7	0.9	0.9	0.4	0.5	0.5	0.4	0.5	0.2	0.2	0.2	0.2	0.3
Belgium	1.0	1.1	1.1	1.1	1.0	0.8	0.8	0.8	1.0	0.9	0.5	0.4	0.4	0.5	0.5	0.3	0.2	0.2	0.3	0.3
United Kingdom	0.9	0.9	0.9	1.0	0.9	0.8	0.7	0.7	0.8	0.8	0.4	0.4	0.3	0.4	0.4	0.2	0.1	0.1	0.2	0.2
United States	0.8	0.9	0.9	1.0	0.9	0.8	0.8	0.8	0.9	0.8	0.4	0.4	0.3	0.3	0.3	0.2	0.2	0.2	0.2	0.2
Japan	1.3	1.1	1.3	1.2	1.2	1.1	1.0	1.0	1.1	1.1	0.6	0.6	0.6	0.6	0.6	0.3	0.3	0.3	0.4	0.4
Average of 9 countries	1.0	1.0	1.0	1.0	1.0	0.9	0.8	0.7	0.9	0.9	0.4	0.4	0.4	0.4	0.4	0.2	0.2	0.2	0.2	0.2

Sources: Consensus Economics, EC, IMF, OECD, The Economist and author's calculations.

3.2. Assessing relative accuracy

Table 2 reports Theil's U coefficient for the comparison of the various institutions' GDP growth forecasts with a same-change naive benchmark. All forecasters have U coefficients lower than one, meaning that they all have a lower MSE than the naive forecast.[21] However, according to the results of the test proposed by Diebold and Mariano (1995), the five forecasters are significantly better than the naive benchmark for current-year but not for next-year horizons. The negative estimates for the parameter α are the equivalent to the result of U coefficients lower than one. For current-year horizons, we are able to reject the null hypothesis of equal forecast accuracy for most countries, at a 10 per cent significance level. For next-year horizons, it is not possible to conclude that the forecasters were significantly better than the naive for the majority of countries (with a clear exception for Japan).

The comparison of the forecast accuracy of the three international organisations with that of the two private institutions is reported in table 3.[22] In general, the $RMSE$ of international organisations' forecasts does not differ much from that of private analysts, for the various countries and horizons. The ratio of $RMSE$ is in most cases close to one. The test of statistical significance of the difference between the two sets of forecasts confirms that, in general, we cannot reject the hypothesis that international organisations and private analysts have similar forecast accuracy. There are just a few cases for the shorter forecasting horizon (Autumn current-year) where this hypothesis is rejected. In most of these cases one of the international organisations, though not always the same, proved to be more accurate than the Consensus or The Economist (ratio of $RMSE$ lower than one $\Leftrightarrow$ negative estimate for α). The evidence is somewhat more consistent for France and Belgium but even for these countries it seems far-fetched to conclude that international organisations perform consistently better in the shorter horizon.[23]

4. Efficiency of forecasts

The evaluation of forecasts provided in the previous section does not assess their quality in the sense of being optimal with regard to a particular information set. To assess this we need to establish testable properties that an optimal forecast should have and, for that, we will assume that the objective function of forecasters is of the mean squared error type, *i.e.* forecasts minimize a symmetric quadratic loss function. As discussed in Timmermann (2007), this implies, under broad conditions, that the optimal forecast is unbiased and there is absence of serial correlation in the forecast errors. The existence of serially correlated errors means that it would be possible to improve the forecast using the information on known past errors. These requirements are usually referred to in the literature as weak efficiency requirements and are empirically tested for our data set.

The test for the weak efficiency requirements is performed directly on the properties of the forecasting errors (unbiasedness and absence of serial correlation). Indeed, for a h-period-ahead forecast to be

---

21 This same-change naive benchmark proved to be more demanding than a no-change benchmark as we expected: Theil's U coefficients are generally higher. There are a few exceptions for year-ahead forecasts for Germany, Italy and Japan, which experienced around zero GDP growth rates during some years of the sample.

22 Recall that, as explained in Section 2, each international organisation is compared with its specific data set for the Consensus and for The Economist.

23 We also run a Diebold and Mariano (1995) test for differences in accuracy among the international organisations and among the two private analysts and, again, it was not possible to reject equal forecast accuracy for the vast majority of cases.

Table 2

GDP GROWTH - COMPARISON OF THE FORECAST ACCURACY OF EACH INSTITUTION WITH THAT OF A SAME-CHANGE NAIVE FORECAST

	Spring next-year forecast					Autumn next-year forecast					Spring current-year forecast					Autumn current-year forecast				
	IMF	EC	OECD	Consen-sus	The Econo-mist	IMF	EC	OECD	Consen-sus	The Econo-mist	IMF	EC	OECD	Consen-sus	The Econo-mist	IMF	EC	OECD	Consen-sus	The Econo-mist
Theil's U																				
Germany	0.9	0.8	0.7	0.8	0.8	0.6	0.4	0.3	0.6	0.5	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0
France	0.9	0.9	0.9	0.9	0.8	0.6	0.4	0.4	0.7	0.6	0.2	0.2	0.2	0.2	0.2	0.0	0.0	0.0	0.1	0.1
Italy	0.8	0.9	0.9	0.9	0.9	0.6	0.5	0.4	0.7	0.7	0.2	0.2	0.2	0.3	0.2	0.1	0.0	0.0	0.1	0.1
Spain	0.5	0.6	0.5	0.6	0.6	0.2	0.3	0.2	0.3	0.3	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0
Netherlands	0.6	0.7	0.7	0.7	0.7	0.5	0.4	0.4	0.5	0.5	0.2	0.2	0.2	0.2	0.2	0.1	0.0	0.0	0.1	0.1
Belgium	0.6	0.7	0.7	0.7	0.7	0.3	0.3	0.3	0.5	0.5	0.2	0.2	0.2	0.2	0.2	0.1	0.0	0.0	0.1	0.1
United Kingdom	0.5	0.5	0.4	0.5	0.5	0.3	0.3	0.3	0.4	0.3	0.1	0.2	0.1	0.2	0.2	0.0	0.0	0.0	0.0	0.0
United States	0.4	0.5	0.5	0.6	0.5	0.4	0.4	0.4	0.5	0.4	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Japan	0.8	0.6	0.8	0.7	0.7	0.5	0.4	0.4	0.5	0.6	0.4	0.4	0.3	0.3	0.3	0.1	0.1	0.1	0.1	0.1
Diebold and Mariano test: estimate for α																				
Germany	-0.5	-1.5	-2.1	-1.4	-1.6	-2.1	-4.6	-5.1	-3.0	-3.4	-4.3	-4.5	-4.2	-4.6	-4.2	-4.4	-4.8	-4.6	-4.7	-4.5
France	-0.4	-0.3	-0.4	-0.5	-0.5	-1.2	-1.8	-2.1	-1.1	-1.2	-1.8	-1.7	-1.8	-1.7	-1.7	-2.0	-2.1	-2.2	-2.1	-2.1
Italy	-1.1	-0.7	-0.7	-0.5	-0.5	-2.1	-2.3	-3.0	-1.5	-1.5	-2.2	-2.3	-2.3	-2.0	-2.1	-2.6	-2.6	-2.6	-2.5	-2.5
Spain	-2.4	-2.7	-3.1	-2.0	-2.0	-4.3	-4.5	-4.8	-4.0	-4.0	-2.6	-2.5	-2.6	-2.5	-2.6	-2.6	-2.5	-2.5	-2.6	-2.5
Netherlands	-2.3	-2.0	-1.9	-1.7	-1.8	-2.9	-3.7	-3.8	-2.8	-2.7	-2.9	-2.7	-2.8	-3.0	-2.9	-3.3	-3.2	-3.2	-3.3	-3.2
Belgium	-1.6	-1.5	-1.5	-1.3	-1.4	-3.0	-3.6	-3.5	-2.2	-2.5	-2.7	-2.8	-2.7	-2.5	-2.7	-3.1	-3.4	-3.1	-2.9	-3.0
United Kingdom	-3.7	-3.7	-3.7	-3.4	-3.4	-4.2	-4.8	-4.8	-4.0	-4.2	-2.8	-2.8	-2.9	-2.7	-2.7	-3.4	-3.4	-3.4	-3.4	-3.4
United States	-2.6	-2.4	-2.3	-1.8	-2.3	-2.7	-2.8	-2.7	-2.2	-2.6	-2.5	-2.3	-2.6	-2.5	-2.5	-2.3	-2.3	-2.3	-2.3	-2.3
Japan	-2.4	-3.8	-2.5	-2.8	-2.8	-4.9	-6.1	-6.1	-5.0	-4.6	-3.2	-3.3	-3.4	-3.4	-3.3	-4.3	-4.5	-4.3	-4.1	-4.1
p-value of the t-statistic for $\alpha=0$[(a)]																				
Germany	0.76	0.31	0.24	0.25	0.24	0.33	0.09	0.08	0.10	0.11	0.07	0.07	0.06	0.06	0.06	0.05	0.04	0.04	0.04	0.04
France	0.56	0.63	0.56	0.49	0.48	0.25	0.06	0.07	0.19	0.16	0.02	0.02	0.04	0.03	0.03	0.01	0.01	0.02	0.02	0.02
Italy	0.42	0.53	0.50	0.56	0.58	0.17	0.09	0.10	0.18	0.18	0.04	0.04	0.04	0.04	0.03	0.02	0.02	0.02	0.03	0.02
Spain	0.24	0.15	0.19	0.26	0.29	0.17	0.10	0.11	0.16	0.17	0.12	0.12	0.11	0.12	0.13	0.10	0.07	0.07	0.10	0.11
Netherlands	0.27	0.25	0.26	0.35	0.32	0.20	0.11	0.15	0.18	0.22	0.20	0.21	0.20	0.19	0.19	0.14	0.11	0.10	0.14	0.15
Belgium	0.23	0.19	0.15	0.10	0.16	0.10	0.04	0.03	0.05	0.08	0.02	0.02	0.01	0.02	0.02	0.01	0.00	0.01	0.02	0.02
United Kingdom	0.11	0.12	0.12	0.12	0.12	0.13	0.14	0.14	0.11	0.13	0.14	0.13	0.14	0.15	0.15	0.08	0.07	0.07	0.07	0.07
United States	0.09	0.11	0.13	0.18	0.09	0.07	0.07	0.11	0.10	0.07	0.02	0.03	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.01
Japan	0.06	0.01	0.04	0.05	0.02	0.01	0.01	0.01	0.01	0.01	0.04	0.03	0.02	0.03	0.03	0.01	0.01	0.01	0.01	0.01

Sources: Consensus Economics, EC, IMF, OECD, The Economist and author's calculations.
Note: (a) p-values below or equal to 0.05 (0.1) are shaded in dark yellow (light yellow) and indicate rejection of the null hypothesis of equal forecast accuracy, at a significance level of 5 (10) per cent.

Table 3

GDP GROWTH – COMPARISON OF THE FORECAST ACCURACY OF INTERNATIONAL ORGANISATIONS WITH THAT OF PRIVATE ANALYSTS

	Compared to the Consensus												Compared to The Economist											
	Spring next-year forecast			Autumn next-year forecast			Spring current-year forecast			Autumn current-year forecast			Spring next-year forecast			Autumn next-year forecast			Spring current-year forecast			Autumn current-year forecast		
	IMF	EC	OECD	IMF	EC	OECD	IMF	EC	OECD	IMF	EC	OECD	IMF	EC	OECD	IMF	EC	OECD	IMF	EC	OECD	IMF	EC	OECD
Ratio of $RMSE$[a]																								
Germany	1.0	1.0	1.0	1.0	1.0	1.0	1.1	1.1	1.2	1.1	0.9	1.1	1.0	1.0	1.0	1.0	1.0	0.9	1.0	1.1	1.1	0.8	0.7	0.9
France	0.9	1.0	1.0	0.9	0.9	0.9	0.9	1.1	1.0	0.6	0.7	1.1	0.9	1.0	1.0	0.9	1.0	0.9	0.9	1.0	1.0	0.6	0.8	1.0
Italy	1.0	1.0	1.0	0.9	1.0	0.9	1.0	0.9	1.0	1.1	1.0	1.0	0.9	1.0	1.0	0.9	1.1	0.9	1.0	0.9	1.0	1.2	1.1	1.0
Spain	0.9	1.0	0.9	0.9	1.0	0.9	0.8	0.9	0.9	1.2	0.6	0.9	0.9	1.0	0.9	0.9	1.1	0.9	0.9	0.9	0.9	0.9	0.6	0.8
Netherlands	0.9	1.0	1.0	1.0	0.9	1.0	1.0	1.1	1.1	1.0	1.1	1.0	0.9	1.0	1.0	1.0	1.0	0.9	0.9	1.1	1.1	0.8	1.1	1.0
Belgium	0.9	1.0	1.0	0.8	0.8	0.9	0.9	0.9	0.9	0.7	0.6	0.7	0.9	1.0	1.0	0.9	1.0	0.9	1.0	1.0	0.9	0.9	0.7	0.8
United Kingdom	1.0	1.0	1.0	1.0	0.9	1.0	0.9	0.9	0.9	1.0	0.8	1.1	1.0	1.0	1.0	1.0	1.0	1.0	0.9	0.9	0.9	1.0	0.8	1.2
United States	0.8	0.9	1.0	0.9	1.0	1.0	1.0	1.4	1.0	0.9	1.0	1.1	0.9	1.0	1.0	1.0	1.0	1.0	1.1	1.3	1.0	0.9	1.0	1.1
Japan	1.0	0.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.8	0.9	1.0	1.0	0.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.8	0.9	1.0
Diebold and Mariano test: estimate for α																								
Germany	-0.2	0.2	-0.2	-0.2	-0.1	-0.1	0.1	0.1	0.2	0.0	-0.0	0.0	-0.3	0.1	-0.2	-0.3	0.1	-0.4	0.1	0.1	0.1	-0.0	-0.1	-0.0
France	-0.4	-0.1	0.0	-0.4	-0.4	-0.2	-0.1	0.0	-0.0	-0.1	-0.1	0.0	-0.3	-0.2	0.1	-0.2	-0.0	-0.2	-0.1	-0.0	-0.0	-0.1	-0.1	-0.0
Italy	-0.4	0.1	0.1	-0.5	0.0	-0.3	-0.0	-0.1	-0.0	0.0	-0.0	-0.0	-0.5	-0.1	-0.1	-0.4	0.4	-0.3	0.0	-0.1	-0.0	0.1	0.0	0.0
Spain	-0.4	-0.1	-0.4	-0.4	-0.1	-0.2	-0.1	-0.1	-0.1	0.0	-0.0	-0.0	-0.4	-0.2	-0.4	-0.4	0.2	-0.2	-0.1	-0.1	-0.1	-0.0	-0.0	-0.0
Netherlands	-0.5	-0.2	-0.1	-0.3	-0.3	-0.0	-0.0	0.1	0.1	-0.0	0.0	0.0	-0.5	-0.2	-0.1	-0.3	-0.0	-0.3	-0.1	0.1	0.1	-0.1	0.0	0.0
Belgium	-0.5	-0.2	-0.1	-0.8	-0.7	-0.5	-0.2	-0.1	-0.1	-0.1	-0.1	-0.1	-0.4	-0.1	0.1	-0.6	-0.0	-0.3	-0.1	-0.0	-0.1	-0.0	-0.1	-0.1
United Kingdom	-0.2	-0.0	-0.0	-0.2	-0.4	-0.0	-0.1	-0.1	-0.0	0.0	-0.0	0.0	-0.1	-0.2	-0.0	-0.0	-0.1	-0.1	-0.1	-0.1	-0.1	0.0	-0.0	0.0
United States	-0.8	-0.5	-0.3	-0.5	-0.1	0.1	0.0	0.2	0.0	-0.0	0.0	0.0	-0.3	-0.2	0.0	-0.1	-0.0	-0.0	0.1	0.2	0.0	-0.0	-0.0	0.0
Japan	0.4	-1.0	0.4	0.1	-0.3	0.3	0.2	0.1	-0.1	-0.2	-0.1	-0.0	0.3	-0.9	0.4	-0.3	-0.1	-0.0	0.1	0.1	-0.1	-0.2	-0.1	0.0
p-value of the t-statistic for $\alpha=0$[b]																								
Germany	0.75	0.34	0.63	0.72	0.85	0.72	0.34	0.19	0.09	0.53	0.34	0.14	0.57	0.62	0.67	0.43	0.50	0.34	0.51	0.47	0.09	0.23	0.05	0.41
France	0.49	0.76	0.84	0.37	0.29	0.08	0.25	0.47	0.92	0.02	0.01	0.34	0.42	0.62	0.33	0.46	0.82	0.34	0.18	0.92	0.51	0.04	0.05	0.82
Italy	0.60	0.84	0.59	0.40	0.96	0.30	0.84	0.26	0.64	0.74	0.75	0.97	0.49	0.81	0.58	0.32	0.05	0.32	0.91	0.55	0.62	0.53	0.35	0.90
Spain	0.22	0.60	0.43	0.28	0.27	0.22	0.18	0.21	0.12	0.04	0.20	0.63	0.09	0.37	0.53	0.18	0.29	0.31	0.36	0.15	0.09	0.38	0.04	0.31
Netherlands	0.07	0.35	0.72	0.17	0.30	0.97	0.71	0.35	0.11	0.89	0.07	0.90	0.09	0.41	0.79	0.16	0.88	0.55	0.32	0.54	0.15	0.10	0.48	0.97
Belgium	0.49	0.69	0.77	0.25	0.22	0.13	0.07	0.38	0.21	0.02	0.00	0.00	0.38	0.81	0.42	0.12	0.89	0.13	0.38	0.76	0.58	0.21	0.01	0.09
United Kingdom	0.47	0.83	0.81	0.65	0.40	0.86	0.27	0.32	0.56	0.75	0.17	0.23	0.38	0.35	0.84	0.96	0.15	0.04	0.28	0.41	0.32	0.93	0.11	0.18
United States	0.35	0.52	0.62	0.38	0.79	0.60	0.26	0.13	0.69	0.07	0.90	0.05	0.50	0.65	0.97	0.75	0.90	0.96	0.09	0.22	0.80	0.16	1.00	0.49
Japan	0.59	0.25	0.47	0.86	0.63	0.43	0.38	0.72	0.56	0.02	0.32	0.86	0.48	0.17	0.33	0.59	0.69	0.95	0.45	0.77	0.63	0.05	0.31	0.95

Sources: Consensus Economics, EC, IMF, OECD, The Economist and author's calculations.

Notes: (a) Ratio of the $RMSE$ of each international organisation to the $RMSE$ of the Consensus or The Economist. **(b)** p-values below or equal to 0.05 (0.1) are shaded in dark yellow (light yellow) and indicate rejection of the null hypothesis of equal forecast accuracy, at a significance level of 5 (10) per cent.

efficient, forecast errors can follow a moving average process of order not higher than $h-1$.[24] To implement the test we estimate the regression:

$$e_{t,h} = \gamma + \beta e_{t-1,h} + \varepsilon_{t,h} \qquad (7)$$

and perform the three following tests: a t-test for $\gamma = 0$ (unbiasedness), a t-test for $\beta = 0$ (no serial correlation) and an F-test for the joint hypothesis $\gamma = 0$ and $\beta = 0$ (weak efficiency). If β is significantly different from zero it would indicate that there is a systematic error with autocorrelation of a higher than appropriate order. For these econometric tests to be valid it must be the case that there is no serial correlation in the residual terms $\varepsilon_{t,h}$. The Breusch-Godfrey test is carried out to test for the presence of serial correlation in the residuals.[25]

The evidence regarding unbiasedness of GDP growth forecasts, presented in table 4, shows that for the majority of countries we are not able to reject that the mean error of year-ahead forecasts is statistically equal to zero. However, as hinted from the analysis in Section 3, forecasters present a tendency to significantly overestimate GDP growth for the major euro area countries in year-ahead horizons.[26] Current-year forecasts have no significant bias for the vast majority of countries and institutions (with a few exceptions for Italy and Spain).[27]

When testing jointly for unbiasedness and no serial correlation of forecast errors, it is not possible in most cases to reject that forecasts are efficient for current-year horizons. For year-ahead horizons, the evidence points to inefficiency of the various institutions' forecasts for some euro area countries. This means that projections could have been improved if either the average bias or the information contained in past errors were properly taken into account.

5. Additional dimensions of forecast accuracy

5.1. Assessing directional accuracy

The traditional quantitative evaluation of macroeconomic forecasts tends to overlook the fact that, even if forecast errors are substantial, forecasts may provide useful information about the qualitative status of an economy, such as the acceleration/deceleration of economic activity. Useful forecasts should go in the right direction. This section investigates the directional accuracy of forecasts, *i.e.* the correctness of the projected direction of change of GDP growth.

---

24 Given that we are working with annual data, we assumed that h could be either equal to 1 (for current-year forecasts) or 2 (for year-ahead forecasts). For $h = 1$, the errors must be serially uncorrelated.

25 In cases deemed necessary, the test for weak efficiency is performed by running an alternative regression: $e_{t,h} = \gamma + \beta_1 e_{t-1,h} + \beta_2 e_{t-2,h} + \varepsilon_{t,h}$ and testing for $\beta_1 = \beta_2 = 0$ (no serial correlation) and for $\gamma = \beta_1 = \beta_2 = 0$ (weak efficiency). Results presented in table 4 for Germany, France, Italy and Spain refer to this equation, given that the Breusch-Godfrey test applied to equation (7) indicated possible serial correlation in the residuals in various cases.

26 The evidence of a significant bias for major euro area countries in year-ahead horizons still holds if we exclude 2009 from the sample.

27 As suggested by Holden and Peel (1990), we also perform a direct test for the statistical significance of the bias by running the regression $e_{t,h} = \gamma + \varepsilon_{t,h}$ and making a simple Student's t-test for $\gamma = 0$. This test confirms in general the results presented in table 4 but there is additional evidence of a significant bias in year-ahead forecasts for Japan, at a 10 per cent significance level. This difference in results is probably related to the above mentioned high standard deviation of forecast errors for Japan.

Table 4

GDP GROWTH – TEST FOR WEAK EFFICIENCY OF FORECASTS

	Spring next-year forecast					Autumn next-year forecast					Spring current-year forecast					Autumn current-year forecast				
	IMF	EC	OECD	Consen-sus	The Econo-mist	IMF	EC	OECD	Consen-sus	The Econo-mist	IMF	EC	OECD	Consen-sus	The Econo-mist	IMF	EC	OECD	Consen-sus	The Econo-mist
Test for unbiasedness ($\gamma=0$)[a]																				
Germany	0.05	0.08	0.08	0.11	0.08	0.10	0.17	0.12	0.16	0.12	0.67	0.95	0.83	0.94	0.74	0.63	0.70	1.00	0.87	0.87
France	0.06	0.05	0.10	0.13	0.10	0.03	0.07	0.10	0.11	0.07	0.33	0.37	0.38	0.26	0.44	0.44	0.49	0.41	0.74	0.76
Italy	0.01	0.01	0.03	0.02	0.02	0.01	0.03	0.03	0.03	0.02	0.02	0.00	0.12	0.02	0.05	0.18	0.14	0.29	0.04	0.13
Spain	0.13	0.20	0.25	0.30	0.25	0.51	0.54	0.74	0.54	0.70	0.56	0.16	0.01	0.58	0.28	0.04	0.28	0.03	0.00	0.01
Netherlands	0.29	0.35	0.32	0.32	0.37	0.25	0.36	0.40	0.28	0.32	0.84	0.82	0.69	0.92	0.93	0.25	0.10	0.51	0.29	0.43
Belgium	0.24	0.12	0.14	0.34	0.34	0.20	0.15	0.17	0.26	0.27	0.30	0.92	0.69	0.93	0.68	0.13	0.57	0.48	0.23	0.25
United Kingdom	0.40	0.37	0.37	0.37	0.37	0.40	0.47	0.55	0.43	0.34	0.69	0.82	0.91	0.85	0.76	0.67	0.73	0.99	0.88	0.96
United States	0.90	0.95	0.85	0.61	0.70	0.68	0.21	0.44	0.85	0.95	0.20	0.21	0.63	0.51	0.43	0.31	0.64	0.96	0.54	0.60
Japan	0.27	0.24	0.24	0.33	0.26	0.19	0.23	0.25	0.27	0.20	0.74	0.41	0.54	0.97	0.80	0.28	0.29	0.38	0.11	0.16
Test for no serial correlation ($\beta=0$)[b]																				
Germany	0.41	0.49	0.59	0.36	0.26	0.72	0.59	0.65	0.77	0.73	0.79	0.94	0.77	0.50	0.88	0.60	0.64	0.75	0.68	0.64
France	0.59	0.72	0.44	0.29	0.34	0.45	0.43	0.52	0.73	0.69	0.12	0.50	0.41	0.15	0.22	0.49	0.10	0.50	0.23	0.39
Italy	0.08	0.10	0.13	0.08	0.08	0.16	0.37	0.30	0.18	0.18	0.09	0.10	0.34	0.36	0.50	0.29	0.55	0.34	0.18	0.29
Spain	0.00	0.05	0.13	0.01	0.01	0.04	0.18	0.24	0.09	0.03	0.52	0.21	0.10	0.44	0.39	0.46	0.17	0.23	0.04	0.09
Netherlands	0.24	0.27	0.37	0.25	0.22	0.36	0.78	0.84	0.31	0.28	0.75	0.26	0.36	0.35	0.34	0.16	0.23	0.40	0.67	0.23
Belgium	0.94	0.71	0.76	0.64	0.55	0.97	0.37	0.55	0.58	0.72	0.05	0.34	0.66	0.17	0.23	0.40	0.73	0.97	0.45	0.52
United Kingdom	0.16	0.15	0.16	0.18	0.19	0.39	0.54	0.68	0.49	0.62	0.72	0.92	0.96	0.80	0.72	0.31	0.26	0.38	0.72	0.91
United States	0.10	0.19	0.11	0.07	0.09	0.34	0.93	0.78	0.23	0.22	0.65	0.19	0.64	0.68	0.71	0.52	0.98	0.79	0.91	0.82
Japan	0.15	0.22	0.31	0.18	0.23	0.24	0.38	0.63	0.25	0.33	0.46	0.38	0.35	0.46	0.34	0.11	0.32	0.41	0.29	0.32
Test for weak efficiency ($\gamma=0$ and $\beta=0$)[c]																				
Germany	0.10	0.18	0.14	0.20	0.11	0.39	0.49	0.46	0.51	0.42	0.87	0.99	0.91	0.70	0.95	0.77	0.76	0.89	0.85	0.82
France	0.04	0.09	0.06	0.08	0.07	0.13	0.28	0.36	0.41	0.29	0.20	0.62	0.54	0.24	0.36	0.65	0.17	0.62	0.38	0.57
Italy	0.00	0.01	0.01	0.01	0.01	0.02	0.07	0.07	0.04	0.03	0.07	0.02	0.32	0.10	0.18	0.37	0.18	0.22	0.13	0.28
Spain	0.00	0.07	0.16	0.02	0.01	0.06	0.28	0.38	0.12	0.04	0.61	0.21	0.04	0.61	0.50	0.15	0.16	0.10	0.03	0.05
Netherlands	0.19	0.27	0.29	0.24	0.25	0.29	0.61	0.66	0.29	0.31	0.93	0.51	0.58	0.63	0.62	0.26	0.19	0.58	0.39	0.20
Belgium	0.43	0.28	0.31	0.47	0.42	0.43	0.31	0.37	0.40	0.47	0.10	0.63	0.83	0.38	0.45	0.27	0.82	0.77	0.42	0.47
United Kingdom	0.17	0.14	0.17	0.23	0.21	0.38	0.57	0.73	0.50	0.46	0.88	0.97	0.99	0.95	0.90	0.56	0.50	0.67	0.92	0.99
United States	0.25	0.36	0.27	0.17	0.22	0.49	0.38	0.64	0.48	0.44	0.26	0.07	0.78	0.70	0.61	0.52	0.89	0.96	0.82	0.85
Japan	0.06	0.10	0.15	0.14	0.11	0.08	0.19	0.36	0.17	0.14	0.74	0.56	0.59	0.75	0.62	0.20	0.43	0.54	0.23	0.30

Sources: Consensus Economics, EC, IMF, OECD, The Economist and author's calculations.

Notes: (a) p-value of the t-statistic for $\gamma=0$. p-values below or equal to 0.05 (0.1) are shaded in dark yellow (light yellow) and indicate rejection of the null hypothesis of unbiasedness, at a significance level of 5 (10) per cent. **(b)** p-value of the t-statistic for $\beta=0$. In the cases of Germany, France, Italy and Spain, p-value of the F-statistic for $\beta_1=\beta_2=0$. p-values below or equal to 0.05 (0.1) are shaded in dark yellow (light yellow) and indicate rejection of the null hypothesis of no serial correlation, at a significance level of 5 (10) per cent. **(c)** p-value of the F-statistic for $\gamma=\beta=0$. In the cases of Germany, France, Italy and Spain, p-value of the F-statistic for $\gamma=\beta_1=\beta_2=0$. p-values below or equal to 0.05 (0.1) are shaded in dark yellow (light yellow) and indicate rejection of the null hypothesis of weak efficiency, at a significance level of 5 (10) per cent.

Being y_t the actual growth rate in year t, let $\Delta y_t = y_t - y_{t-1}$ be the actual acceleration $\left(\Delta y_t > 0\right)$ or deceleration $\left(\Delta y_t < 0\right)$ in year t. Most previous studies compute the predicted acceleration/deceleration by comparing the forecasted growth rate with the actual growth rate of the previous period $\left(\hat{\Delta y}_{t,h} = \hat{y}_{t,h} - y_{t-1}\right)$. However, for longer forecasting horizons this would imply using information not yet known to forecasters at the time of forecasting. To be consistent with the approach followed in Section 3 – use only information available to forecasters at each point in time – and following the methodology of Ashiya (2003), we decided to compute the predicted direction of change as the acceleration/deceleration implicit in the forecast at each forecasting exercise $\left(\hat{\hat{\Delta y}}_{t,h} = \hat{y}_{t,h} - \hat{y}_{t-1,h}\right)$. To evaluate the directional accuracy of forecasts the sign of $\hat{\hat{\Delta y}}_{t,h}$ is compared to the sign of Δy_t.

The directional data for each country can be arranged in a 2x2 contingency table, in which the two rows represent positive and negative/null changes in the outcome and the two columns represent positive and negative/null changes in the forecast. If the number of cases in the diagonal ($n_{11} + n_{22} =$ cases where Δy_t and $\hat{\hat{\Delta y}}_{t,h}$ are both > 0 or both ≤ 0) is "sufficiently" large compared to the total number of observations $\left(n\right)$, the forecasts are considered to be directionally accurate. More formally, we run a chi-squared independence test as described in Carnot *et al.* (2005):[28]

$$\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left(n_{ij} - n_{i.}n_{.j}/n\right)^2}{n_{i.}n_{.j}/n} \sim \chi^2(1) \qquad (8)$$

The null hypothesis is that the sign of Δy_t and the sign of $\hat{\hat{\Delta y}}_{t,h}$ are independent. The rejection of the null means that there is a significant association between the actual and the predicted direction of change and, therefore, forecasts can be said to be directionally accurate.

As before, the directional accuracy of the various forecasters is compared to that of a same-sign of change naive benchmark. This naive benchmark extrapolates the same sign of change for GDP growth as was last observed at the time of forecasting. Also, the forecasting ability of the three international organisations in terms of direction of change is compared to that of the two private sector institutions.

Table 5 shows the proportion of times that forecasters correctly predicted that GDP was going to accelerate or decelerate. For the group of nine countries, forecasts of all institutions are accurate more than 60/70 per cent of the time for the year-ahead horizons. For current-year horizons their accuracy is higher, at around 80/90 per cent of the time.[29] The results of the chi-squared independence test for the individual countries confirm that there is a significant association between the sign of change of GDP growth in the forecasts and in the outcomes for basically all countries, with some exceptions for the longest forecasting horizon.

When looking at different benchmarks to evaluate the directional accuracy of forecasts, it is clear that the five forecasters were better at predicting the sign of change of GDP growth than a naive forecast for all

28 See Ash *et al.* (1998) for an application of alternative non-parametric tests on the direction of forecasts.

29 Note that, for this group of countries, the sign of $\hat{\hat{\Delta y}}_{t,h}$ proved to be a more accurate predictor than the sign of $\hat{\Delta y}_{t,h}$ for year-ahead horizons. This is in line with previous results by Ashiya (2003).

Table 5

GDP GROWTH – DIRECTIONAL ACCURACY OF FORECASTS

	Spring next-year forecast					Autumn next-year forecast					Spring current-year forecast					Autumn current-year forecast				
	IMF	EC	OECD	Consen-sus	The Econo-mist	IMF	EC	OECD	Consen-sus	The Econo-mist	IMF	EC	OECD	Consen-sus	The Econo-mist	IMF	EC	OECD	Consen-sus	The Econo-mist
Percentage of correct predictions of the direction of change																				
Germany	47	71	65	47	71	67	83	83	72	67	83	83	89	89	83	100	100	94	100	94
France	59	65	59	53	71	72	78	78	67	67	94	89	83	89	89	100	94	94	89	89
Italy	53	53	53	47	59	61	72	72	56	56	78	78	78	72	78	94	94	94	89	89
Spain	87	71	82	80	80	88	78	83	81	81	81	78	89	88	88	94	100	100	100	88
Netherlands	67	65	71	60	60	56	72	67	63	56	63	72	72	65	59	81	83	83	76	71
Belgium	73	71	76	80	80	88	89	78	81	88	88	83	83	82	82	88	89	89	76	76
United Kingdom	82	82	76	82	82	83	89	94	89	89	100	94	94	94	94	94	100	100	100	100
United States	59	65	65	59	59	67	56	67	67	67	78	78	83	78	78	83	83	83	78	72
Japan	53	82	59	53	59	78	83	67	78	72	72	78	72	72	72	72	78	72	78	72
All 9 countries	64	69	67	62	69	73	78	77	72	71	82	81	83	81	81	90	91	90	87	84
Ratio of correct predictions to those of a naive benchmark																				
Germany	0.8	1.2	1.2	0.9	1.5	1.1	1.6	1.8	1.4	1.6	2.1	2.1	2.0	2.0	1.7	2.6	2.3	1.9	2.3	1.7
France	1.0	1.1	1.0	0.9	1.2	1.4	1.5	1.3	1.1	1.1	2.8	2.7	2.1	2.3	2.3	3.6	3.4	2.4	2.3	2.3
Italy	1.1	1.1	1.1	1.0	1.3	1.3	1.5	1.5	1.2	1.2	1.6	1.6	1.6	1.4	1.6	1.9	1.9	1.9	1.8	1.8
Spain	2.0	1.7	2.0	1.9	2.3	2.2	1.9	2.0	2.0	2.3	1.6	1.4	1.6	1.8	1.8	1.9	1.8	1.8	2.0	1.8
Netherlands	1.3	1.0	1.1	0.9	1.1	0.9	1.4	1.3	1.2	1.2	1.7	1.6	1.6	1.7	1.8	1.9	1.5	1.5	1.5	1.6
Belgium	2.6	2.0	2.2	2.8	3.4	2.6	2.5	2.2	2.4	3.0	2.0	1.7	1.7	1.9	1.9	2.0	1.8	1.8	1.7	1.7
United Kingdom	3.5	3.5	3.3	3.5	3.5	3.5	5.0	4.0	3.8	3.8	2.3	2.1	2.1	2.1	2.1	2.1	2.0	2.3	2.3	2.3
United States	1.3	1.2	1.2	1.1	1.1	1.4	0.9	1.1	1.1	1.1	1.4	1.6	1.7	1.6	1.6	1.3	1.3	1.3	1.2	1.1
Japan	1.3	2.3	1.4	1.3	1.4	1.9	2.4	1.6	1.9	1.8	1.1	1.3	1.1	1.1	1.1	1.1	1.3	1.1	1.2	1.1
All 9 countries	1.4	1.5	1.5	1.4	1.6	1.6	1.8	1.7	1.6	1.7	1.8	1.7	1.7	1.7	1.7	1.9	1.8	1.7	1.7	1.6
p-value of the χ^2 statistic[(a)]																				
Germany	0.27	0.02	0.06	0.27	0.03	0.03	0.00	0.00	0.02	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
France	0.09	0.05	0.09	0.16	0.02	0.02	0.01	0.01	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Italy	0.16	0.16	0.16	0.27	0.09	0.06	0.02	0.02	0.11	0.11	0.01	0.01	0.01	0.02	0.01	0.00	0.00	0.00	0.00	0.00
Spain	0.00	0.08	0.01	0.02	0.02	0.00	0.02	0.00	0.01	0.01	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Netherlands	0.19	0.20	0.06	0.45	0.45	0.61	0.06	0.17	0.30	0.61	0.30	0.06	0.06	0.23	0.46	0.01	0.00	0.00	0.02	0.03
Belgium	0.06	0.03	0.01	0.01	0.01	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.03	0.03
United Kingdom	0.01	0.01	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
United States	0.38	0.13	0.13	0.38	0.38	0.04	0.13	0.04	0.09	0.09	0.01	0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.02	0.04
Japan	0.60	0.00	0.31	0.60	0.31	0.01	0.00	0.17	0.02	0.06	0.06	0.02	0.06	0.06	0.06	0.07	0.02	0.07	0.02	0.07

Sources: Consensus Economics, EC, IMF, OECD, The Economist and author's calculations.

Note: (a) p-values below or equal to 0.05 (0.1) are shaded in dark yellow (light yellow) and indicate rejection of the null hypothesis of independence, at a significance level of 5 (10) per cent.

horizons, even if less so for the longest one.[30] When we compare the institutions among themselves,[31] the directional accuracy of international organisations' forecasts does not seem in general to differ significantly from that of the Consensus or The Economist, for the various horizons.

5.2. Ability to forecast recessions

An additional informative criterion to evaluate macroeconomic forecasts is the ability to predict turning points, considering both the number of actual turns that are correctly predicted and the number of false turns that are predicted. To analyse the forecasters ability to predict economic recessions, we define recession as any year in which real GDP declined $\left(y_t < 0\right)$.[32]

Over the sample period 1991-2009, a total of twenty-three recession episodes were identified for the group of nine countries under analysis. The properties of forecasts during those recession episodes are summarised in table 6. When we compute the percentage of episodes that forecasters were able to anticipate, we see that in general they are not able to anticipate in the preceding year that a recession is going to occur. This is particularly true as of Spring of the previous year and more evident in the case of private analysts. Forecasters seem to identify recessions just in the year in which they occur, though by Spring of that year around half of the recession episodes are still not acknowledged by most forecasters. By Autumn of the year of the recession, even though the decline in GDP is correctly identified in the vast majority of cases, the magnitude of the fall is still under-predicted for around 50 per cent of the cases.[33]

During the period analysed, forecasters predicted a couple of false recessions. This is however a rare event and in most cases happened in current-year forecasts for years with close to zero GDP growth outcomes.

The evidence on the difficulties that forecasters experience in identifying economic recessions in advance (or even when they are occurring) is notable, both for international organisations and private analysts. Though the reasons for this do not seem to have been yet adequately explored, some authors such as Loungani (2001) have suggested that either forecasters lack the required information (reliable real-time data or models) or lack the incentives to predict recessions. In any case, we should keep in mind that these point forecasts reported by the various institutions may not capture shifts in the probability that they attach to worst case scenarios.

6. General summary and comparison with previous evaluations

In this article, we assessed the accuracy of IMF's, EC's and OECD's forecasts and compared it with that of the Consensus' and The Economist's surveys of private analysts. The focus was on economic growth forecasts for nine advanced economies, over the past two decades. We now provide an overall picture of our findings and briefly compare them with previous results from in-house evaluations of international organisations' forecasts.

---

30 When we apply a chi-squared independence test to the naive benchmark it is not possible in general to reject the null hypothesis of no significant association between the actual direction of change of GDP growth and that of the naive forecast.

31 Looking at the ratio of correct predictions of each international organisation to those of its corresponding data set for the Consensus and for The Economist (not provided in table 5).

32 A similar analysis of Consensus' forecasts for a large group of countries can be found in Loungani (2001).

33 As mentioned in Section 3, forecasters show a tendency to overestimate growth when the economy is slowing down and this is particularly severe during economic recessions.

Table 6

GDP GROWTH – FORECAST PERFORMANCE DURING RECESSION EPISODES IN THE GROUP OF 9 COUNTRIES OVER THE PERIOD 1991-2009

	Spring next-year forecast					Autumn next-year forecast					Spring current-year forecast					Autumn current-year forecast				
	IMF	EC	OECD	Consen-sus	The Econo-mist	IMF	EC	OECD	Consen-sus	The Econo-mist	IMF	EC	OECD	Consen-sus	The Econo-mist	IMF	EC	OECD	Consen-sus	The Econo-mist
Percentage of episodes where a recession was forecasted $\left(\hat{y}_{t,h} < 0\right)$	0	0	0	0	0	13	22	39	0	0	43	57	65	48	57	87	87	91	83	87
Percentage of episodes where the forecast was too optimistic $\left(\hat{y}_{t,h} > y_t\right)$	100	100	100	100	100	100	100	100	100	100	70	70	61	87	83	48	48	52	61	57
Number of episodes where a false recession was forecasted $\left(\hat{y}_{t,h} < 0, y_t \geq 0\right)$	0	0	0	0	0	0	1	1	0	0	2	3	4	2	3	3	3	3	3	3

Sources: Consensus Economics, EC, IMF, OECD, The Economist and author's calculations.

We find that the accuracy of GDP growth projections clearly increases as the horizon shortens and more information becomes available to the forecaster. Regarding year-ahead horizons, even though the projections of the various forecasters are unbiased and efficient in most cases, there is evidence of inefficiency for some euro area countries. Year-ahead forecasts show a significant negative bias for major euro area countries. This stems from a tendency of the various forecasters to persistently over-predict growth when the economy is slowing down, most noticeably during periods of economic recession. Current-year GDP growth forecasts are generally unbiased and efficient.

Our analysis suggests that the quantitative accuracy of the GDP growth forecasts published by the IMF, the EC and the OECD is not statistically different from that of the Consensus or The Economist, for the various countries and horizons examined. In the rare exceptions observed for the shorter horizon (Autumn current-year), no institution proved to perform consistently better. All five forecasters beat in general a naive model, which projects a GDP growth rate equal to the last one observed, for current-year but not for year-ahead horizons.[34]

Notwithstanding a few distinctive features of the analysis undertaken, our findings are broadly in line with those of the latest in-house assessments of forecasts published by the IMF, the EC and the OECD.[35] Timmermann (2007) analysis of IMF's forecasts, over the period 1990-2003, found that GDP growth forecasts display a tendency for over-prediction in next-year horizons for various advanced economies. However, there is very little evidence on biases or serial correlation of errors for current-year forecasts. IMF's performance is overall statistically similar to that of Consensus, even if the IMF performs slightly better in a few cases for current-year horizons. According to Melander *et al.* (2007) assessment of the EC's forecasts, for the period 1969-2005, growth forecasts for the European Union generally proved to be unbiased and efficient, though there is evidence of the contrary for some Member States. They also concluded that the track record of the EC's forecasts is broadly comparable with the ones of the Consensus, the IMF and the OECD. The review of OECD's growth projections for the G7 countries over the period 1991-2006, carried out by Vogel (2007), found that year-ahead forecasts are less accurate and have a tendency to overestimate the outcome. Current-year projections are, however, unbiased and efficient. The author argues that OECD's forecasts tend to outperform Consensus for the current-year horizon.

Regarding the directional accuracy of GDP growth forecasts, we find that all forecasters are directionally accurate in the various horizons, with some exceptions for the longest one. As before, the directional accuracy of international organisations' forecasts does not seem to differ much from that of private analysts. The five forecasters are better at forecasting accelerations/decelerations of economic activity than a naive benchmark.

There is a general agreement in the literature about the failure of most forecasters to predict economic recessions in advance and, sometimes, to detect them contemporaneously.[36] Notwithstanding the limited number of observations, our brief evaluation of the recession episodes occurred in the sample

---

34 Similarly, the assessment of inflation forecasts performed in Abreu (2011) (which only covers the IMF, the Consensus and The Economist) also concludes that the quantitative accuracy of IMF's forecasts is similar to that of the Consensus and The Economist. The accuracy of these three forecasters is not in most cases statistically different from that of a naive random-walk model. Inflation forecasts are generally unbiased and efficient, both for year-ahead and current-year horizons, even though forecasters also display some tendency to over-predict (under-predict) inflation when it is falling (rising). Also, these three forecasters seem to be slightly more accurate at predicting inflation than GDP growth for year-ahead horizons.

35 For earlier assessments see, for example, Artis (1997), Keereman (1999) and Koutsogeorgopoulou (2000).

36 See Fildes and Stekler (2002) for a survey and Loungani (2001) for evidence across a large sample of industrialised and developing countries.

of nine countries during the period 1991-2009 is totally consistent with this finding. As of Spring of the previous year no forecaster is able to predict that GDP is going to fall and by Spring of the recession year around half of the recession episodes is still not acknowledged by most forecasters. Moreover, the forecasts made in Autumn of the recession year still underestimate its magnitude in around 50 per cent of the cases. This underestimation was particularly notorious during the latest economic recession for all five forecasters. Also, forecasters make very few predictions of recessions that do not occur. As pointed out by McNees (1992), this disturbing evidence about the inability to forecast economic recessions advises the forecast user not to ignore the forecasts but rather to think carefully about plausible outcomes far from the central scenarios.

The findings of this study are in line with previous evidence that current-year forecasts for economic growth in advanced economies present in general desirable features but year-ahead forecasts present a more mixed picture in terms of quantitative and qualitative accuracy. This understanding of how large forecast errors are likely to be and how often forecasters are likely to miss the direction where the economy is going is absolutely necessary in order to assess the usefulness of forecasts to its users. Some may consider disappointing the fact that the performance of reputed international organisations is generally similar to that of panels of private analysts. However, we must emphasize that international organisations' forecasts serve a quite different purpose from those of private institutions. They do provide more than just point forecasts. In particular, they provide a detailed and consistent picture for the international outlook and a thorough discussion of the main issues and risks, besides policy recommendations potentially valuable to policymakers. For the forecast user it might however be comforting to learn that he can place as much (little) confidence in the alternative private analysts' forecasts that are available on a monthly basis.

References

Abreu, I. (2011), "International organisations' vs. private analysts' forecasts: an evaluation", *Working Papers* 20, Banco de Portugal.

Artis, M. J. (1997), "How accurate are the IMF's short-term forecasts? Another examination of the World Economic Outlook", *Staff Studies for the World Economic Outlook – World Economic and Financial Surveys*, International Monetary Fund.

Ash, J. C. K., Smyth, D. J. and Heravi, S. M. (1998), "Are OECD forecasts rational and useful?: a directional analysis", *International Journal of Forecasting* 14(3), 381–391.

Ashiya, M. (2003), "The directional accuracy of 15-months-ahead forecasts made by the IMF", *Applied Economics Letters* 10(6), 331–333.

Carnot, N., Koen, V. and Tissot, B. (2005), *Economic Forecasting*, Palgrave MacMillan.

Diebold, F. X. and Mariano, R. S. (1995), "Comparing predictive accuracy", *Journal of Business and Economic Statistics* 13(3), 253–263.

Fildes, R. and Stekler, H. (2002), "The state of macroeconomic forecasting", *Journal of Macroeconomics* 24(4), 435–468.

Harvey, D., Leybourne, S. and Newbold, P. (1997), "Testing the equality of prediction mean squared errors", *International Journal of Forecasting* 13(2), 281–291.

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

Keereman, F. (1999), "The track record of the Commission forecasts", *Economic Papers* 137, European Commission (Directorate General for Economic and Financial Affairs).

Koutsogeorgopoulou, V. (2000), "A post-mortem on Economic Outlook projections", OECD Economics Department, *Working Papers* 274, OECD.

Lenain, P. (2001), "What is the track record of OECD economic projections?", T*echnical report*, OECD Economics Department.

Loungani, P. (2001), "How accurate are private sector forecasts? Cross-country evidence from consensus forecasts of output growth", *International Journal of Forecasting* 17(3), 419–432.

Loungani, P., Stekler, H. and Tamirisa, N. (2011), "Information rigidity in growth forecasts: Some cross-country evidence", IMF *Working Paper* 125, International Monetary Fund.

McNees, S. (1992), "How large are economic forecast errors?", *New England Economic Review*, pp. 25–42.

Melander, A., Sismanidis, G. and Grenouilleau, D. (2007), "The track record of the Commission's forecasts - an update", *Economic Papers* 291, European Commission (Directorate General for Economic and Financial Affairs).

Theil, H. (1971), *Applied Economic Forecasting*, North-Holland Publishing Company.

Timmermann, A. (2007), "An evaluation of the World Economic Outlook forecasts", *IMF Staff Papers* 54(1), 1–33.

Vogel, L. (2007), "How do the OECD growth projections for the G7 economies perform? A post-mortem", OECD Economics Department *Working Papers* 573, OECD.

Vuchelen, J. and Gutierrez, M.-I. (2005), "Do the OECD 24 month horizon growth forecasts for the G7-countries contain information?", *Applied Economics* 37(8), 855–862.

Zarnowitz, V. and Braun, P. A. (1993), "Twenty-two years of the NBER-ASA quarterly economic outlook surveys: Aspects and comparisons of forecasting performance", *in* J. Stock and M. Watson, eds, "Business Cycles, Indicators and Forecasting", The University of Chicago Press, pp. 11–93.