JEL Classification: E37; E58
Keywords: forecasting accuracy; prediction process; survey forecast

Assessment of Consensus Forecasts Accuracy: The Czech National Bank Perspective*

Filip NOVOTNÝ (filip.novotny@cnb.cz) – *corresponding author*
Marie RAKOVÁ
both authors: Czech National Bank, Prague

Abstract
This paper compares the accuracy of the Consensus forecasts for euro-area GDP growth, consumer and producer price inflation, and the USD/EUR exchange rate to those of the European Commission, International Monetary Fund, and Organization for Economic Co-operation and Development, and also to the naive forecast and the forecast implied by the forward exchange rate. In the period from 1994 to 2009 the Consensus forecasts for effective euro-area consumer price inflation and GDP growth beat the alternatives by a difference which is typically statistically significant. The results are more diverse for the pre-crisis sample (1994–2007). The Consensus forecast for euro-area producer price inflation significantly outperforms the naive forecast in the short term. Finally, the Consensus forecast for the USD/EUR exchange rate during the period from 2002 to 2009 is more precise than the naive forecast and the forecast implied by the forward rate.

1. Introduction

The monetary policy regime of direct inflation targeting was adopted by the Czech National Bank (CNB) in 1998. Under inflation targeting, the forecast for consumer price inflation (CPI) at the monetary policy horizon is of great relevance to the decision-making on current interest rates. An important assumption of the CNB's macroeconomic forecast is the external economic outlook, which is partly derived from the Consensus Forecasts (hereinafter "Consensus").

Consensus is a regular monthly survey publication which provides forecasts and views on the principal macroeconomic indicators, including GDP growth, price inflation, interest rates, and exchange rates, in over 85 countries. The surveys mainly contain the expectations of the main investment banks and other well-known analytical centers. Consensus serves as a background for the decision-making processes of economic agents, including central banks. It is therefore relevant to research the ex-post accuracy of Consensus, because if we are aware of the historical performance of the Consensus forecasts we are then in a better position to judge the information it provides for our future decision-making.

In this paper, we decided to compare the forecasting accuracy of Consensus with the corresponding forecasts of the International Monetary Fund (IMF), Organization for Economic Co-operation and Development (OECD), and European Commission (EC), and also with the naive forecast and forecasts implied by the forward

* We have benefited from valuable comments from the Monetary and Statistics Department of the Czech National Bank on previous versions of this paper as well as comments by the referees of the paper and participants at the 2010 Czech Economic Society Conference. In addition, we are grateful to David Kocourek for his technical assistance. Nevertheless, all errors and omissions are ours. The views expressed in this paper are those of the authors and do not necessarily represent those of the Czech National Bank.

exchange rate. The forecasts of real GDP growth, CPI inflation, and PPI inflation for Austria, France, Germany, and Italy are assessed along with the nominal USD/EUR exchange rate forecasts. Contrary to the available literature, we assess the forecasts for each macroeconomic variable in terms of the effective indicator, where each country forecast is weighted by the respective country share in Czech exports. The same effective indicators are applied in the CNB's prediction process.

Our main focus is on the Consensus performance in the period from 1994 to 2009, but we also separately assess the pre-crisis period from 1994 to 2007 to reveal any potential impacts which the financial crisis might have imposed on the overall forecasting accuracy. In addition, we assess the accuracy of the forecasts for a shorter period starting in 2002. Standard descriptive statistics of the accuracy of the forecasts as well as statistical tests for significance in the forecasting errors are applied.

Because of the rather long history and wide acceptance of Consensus, there are a vast number of publications dealing with its accuracy. Similarly to our paper, Batchelor (2001, 2007) compares the Consensus forecasts with the corresponding forecasts published by the IMF and OECD. According to him, the Consensus forecasts are less biased and more accurate for the G7 countries. The GDP growth forecasts have been overestimated (higher than actual values) in the cases of France, Germany, Italy, and Japan. Conversely, the CPI inflation forecasts have been unbiased. Similarly, Osterloh (2008), who dealt solely with the Consensus forecasts for German real GDP growth in the period from 1995 to 2005, shows that forecasters pooled by Consensus were systematically overestimating the growth rates. In addition, he discovers a relatively low accuracy for next-year forecasts compared to a simple naive forecast. The Consensus forecasts for 12 industrial countries over the period from 1996 to 2006 were also explored regarding their bias and information efficiency in Ager et al. (2009). The authors showed that the forecasts for some countries, e.g., for Germany and Italy, and in particular the forecasts for horizons longer than one year, were systematically biased. Moreover, forecast information efficiency had to be rejected in almost all cases.

Among others, Timmermann (2006) analyzed the IMF forecasts in comparison with Consensus. He finds the IMF forecasts similar to those of Consensus and identifies weaknesses in the IMF forecasts. Bowle et al. (2007) assess the Survey of Professional Forecasters (SPF) in the period from 1999 to 2006. The SPF surveys approximately 75 forecasters from the European Union and asks them for their short and long-term expectations about euro-area economic variables – inflation, GDP growth, and unemployment. Bowle et al. (2007) conclude that the SPF provides better forecasts for the GDP growth of the euro area in comparison with Consensus. Nevertheless, their results are not surprising given that the SPF focuses solely on the euro area and has the advantage of a larger sample of respondents (some of whom contribute to Consensus as well).

A different approach was applied by Dovern et al. (2009). Contrary to the previous literature on the topic, the authors did not concentrate on the mean Consensus, but analyzed the dispersion (heterogeneity) of individual Consensus forecasters. They found that disagreement among forecasters tends to rise during recessions and is particularly pronounced in the case of real variables (GDP growth, consumption growth, investment growth, and the unemployment rate). In addition, there is a down-

ward trend in the disagreement of the forecasts for nominal variables (CPI inflation and interest rates) and this is lower in the case of countries with an independent central bank (for 35 percent).[1] It is also mentioned that Consensus is sensitive to current conditions, due to a strong correlation of the one-year-ahead forecasts with the current actual values.

More generally, Ang et al. (2007) compared four different methods of inflation forecasting and they assert that survey forecasts (the Livingston, Michigan, and SPF surveys) outperform the other three methods, namely, the ARIMA model, the economic model of the Phillips curve, and the term structure of interest rates.

With respect to the Czech National Bank's own inflation forecasts, Antal et al. (2008) analyzed the bias of the forecasts in relation to any undershooting of the inflation target. In addition, Babecký and Podpiera (2011) compared the accuracy of the CNB's inflation forecasts with other financial institutions' forecasts for the Czech Republic.

This paper is organized as follows. The next section provides the basic characteristics of Consensus and deals with the data and methodology description. Section 3 covers the empirical results for three macroeconomic variables: CPI inflation, PPI inflation, and real GDP growth. Section 4 proceeds with the results for the USD/EUR exchange rate, and Section 5 summarizes the main findings.

2. Data and Methodology

Consensus is a regular monthly publication of the London-based Consensus Economics, which was founded in 1989. Consensus Economics pools more than 700 economic analysts and economic research centers, mostly from private investment banks. The number of participating analytical centers varies across individual countries and also for different time periods for a given country. It oscillates mostly between 10 and 30 analytical centers for a certain economic variable and country.

The final Consensus forecast for a given country and a certain economic variable is a simple average of the forecasts provided by each participating forecaster. The general advantage of the mean forecast is that it eliminates (to a certain degree) any possible systematic errors in individual forecasts.

The Consensus forecasts for selected external economic variables started to be used at the Czech National Bank in the second half of 2002. This was connected with the introduction of the bank's unconditional quarterly projection model (QPM) of the Czech economy, which was replaced by the new dynamic structural model (g3) in 2008, and with the resulting need for a consistent forecast of external (foreign) economic variables.

The bank's external economic outlooks use only a limited number of the economic variables which are provided by Consensus. Moreover, the original Consensus forecasts are adjusted to make them appropriate for the bank's prediction process. Specifically, the forecasts for CPI and PPI inflation and GDP growth, which are available as whole-year percentage changes, are equally decomposed into individual quarters in order to make them suitable for the quarterly model. Subsequently,

[1] Countries with an independent central bank during the whole period under review (1989–2006) comprise Canada, France, Germany, and the United States. Countries which did not have an independent central bank for the whole period are Japan and the United Kingdom.

the effective euro-area indicator is calculated by weighting each euro-area country by its share in total Czech exports.

The effective euro-area indicator for the Czech Republic comprises 14 euro--area countries, with only Luxemburg and Malta not being included. Germany (47 percent), Slovakia (14 percent), and France (8 percent) are characterized by having the largest shares in the effective indicator.

The adjustment procedure is easier in the case of the USD/EUR exchange rate, where the Consensus point forecasts (rolling-event forecast) three months and one year ahead are interpolated.

In parallel with Consensus the remaining external economic outlooks for the oil price, gasoline price, and three-month Euribor are derived from market instruments (derivatives). The market implied outlooks are prepared as of the Consensus survey date, i.e., on the second Monday of each calendar month, in order to preserve the time consistency of all the predicted variables. Finally, the outlooks for each variable are put together.

The forecasting accuracy of the Consensus forecasts for real GDP growth and CPI inflation is compared with corresponding forecasts published by international institutions – the EC, the IMF, and the OECD. In addition, the Consensus forecasts are assessed against the naive forecasts. The forecasts for PPI inflation and the USD/ /EUR exchange rate are assessed, due to a lack of alternative forecasts, against the naive forecast only; moreover, the forecast for USD/EUR is assessed also against the forecast derived from the forward exchange rate. The forecast horizon for CPI and PPI inflation and real GDP growth is the current year and the next year. The forecast for the USD/EUR exchange rate is available at three-month, one-year, and two-year horizons.

Our naive forecast is a random walk, AR(1) process, with the coefficient 1. The naive forecast is thus the last-year growth rate of GDP, CPI inflation, and PPI inflation, which is simply prolonged to the current year and the next year. We proceed similarly with the naive forecast for the USD/EUR exchange rate. The actual value of the exchange rate as of the Consensus survey date is used as a constant naive forecast at the three-month, one-year, and two-year horizons. The naive forecasts used in the paper simulate the behavior of a naive forecaster who mechanically projects (without any additional judgment) the last available data into the future. Alternatively, ARIMA forecasts, which are another commonly used benchmark based on past values, would probably provide better forecasts. We therefore deem our benchmark as having the worst available expectation about the future, which is our objective.

We encountered the problem of different publication dates of the compared forecasts. The main difference between Consensus and the international institutions' forecasts was the frequency at which the forecasts are published. The advantage of Consensus is its monthly frequency, while the EC, IMF, and OECD publish their standard forecasts twice a year. The EC and the OECD publish their forecasts every May and November, whereas the IMF forecasts are published one month ahead, i.e., in April and October. Nevertheless, in 2007 the international institutions started to publish interim forecasts as well and the forecasting frequency thus doubled.

In order to provide an evaluation of the Consensus and international institutions' forecasts, only the Consensus issues corresponding to the month of the inter-

national institutions forecasts' releases were used for relative forecast assessment. Accordingly, the IMF forecasts were compared with the April and October Consensus issues, and the EC and OECD forecasts with the May and November Consensus issues.

In addition, the real GDP growth and CPI and PPI inflation forecasts were assessed in terms of the effective indicator and not as a single forecast for a given country. This approach reflects the procedure used by the CNB whereby external economic developments are proxied by effective indicators where each country variable is weighted by the country share in total Czech exports. The comparable effective Consensus, EC, IMF, and OECD forecasts are constructed before the forecasts are evaluated.

The effective indicator forecast can be expressed as follows:

$$f_h^{eff} = \sum_{i=1}^{n} f_h^i w_i \tag{1}$$

where f_h^{eff} is the effective indicator forecast at horizon h ($h = 1$ for the current year and $h = 2$ for the next year), f_h^i is a particular country forecast at horizon h, and w_i is the country share in Czech exports $\left(\sum_{i=1}^{n} w_i = 1 \right)$.

Only Austria, France, Germany, and Italy are covered in our effective indicator. These four countries are used because previous forecasts are not available for every euro-area country. Nevertheless, this is a sufficient approximation given that Austria, France, Germany, and Italy account for approximately 70 percent of total Czech exports.

The weight of Germany in our restricted effective indicator is 67 percent. France is weighted by 12 percent, and Italy and Austria by 11 and 10 percent, respectively. We calculate both the effective actual time series of real GDP growth and CPI and PPI inflation, and also their corresponding effective forecasts for each institution, including the naive forecasts.

The forecasts are assessed for the period from 1994 to 2009, with forecast accuracy also being assessed for the shorter period from 2002 to 2009. The motivation for the shorter period is that Consensus was implemented at the CNB in 2002. The forecast for the USD/EUR exchange rate is assessed for the periods from 1999 to 2009 and from 2002 to 2009. In addition, the forecasts are assessed separately in the pre-crisis period (excluding 2008 and 2009) to reveal the potential effects of the financial crisis on overall forecasting accuracy (see the *Appendix* – available on the web-page of this journal: http://journal.fsv.cuni.cz).

In order to assess forecasting accuracy, the forecasting errors are calculated first:

$$e_t = a_t - f_t \tag{2}$$

where a_t represents the actual (realized) value and f_t is its corresponding forecast value. Similarly to Batchelor (2001) and Osterloh (2008), we use unrevised actual values by taking the actual value of the previous year from the current year Consensus June issue. The actual value of GDP growth in 1994 is thus taken from the June 1995 issue of Consensus. We proceed uniformly in the following years.

Unrevised actual data is used because economic forecasters are assumed to be quite unlikely to anticipate the extent of data revisions. It is more likely that forecasters will make their forecasts anticipating the same methods of data construction as those used by governmental statistical agencies. Our approach is the opposite to that used, for example, by Croushore (2010), who based his analysis solely on the latest available time series of actual data, which naturally contain *ex post* revisions.

If we look at the actual time series of effective GDP growth, which is probably the most affected by data revisions, we find that the difference between our method and the actual time series downloaded at the end of 2010 is, on average, zero. Nevertheless, there are differences in individual years. The largest negative deviation (-0.6 percentage points) was observed in 2006 and the largest positive deviation (0.5 percentage points) in 2004. The standard deviation over the whole sample is 0.3 percentage points.

The following standard measures are considered for the descriptive analysis of forecasting accuracy: *mean forecast error* (MFE), *mean absolute forecast error* (MAFE), *mean absolute percentage error* (MAPE), which, in addition to MAFE, represents the forecast errors as a percentage of the actual values, *mean squared error* (MSE), and *root mean squared error* (RMSE).

A forecast is considered to be biased if it is systematically too high or too low. As in, e.g., Ager et al. (2009) and Croushore (2010), we verify the forecast bias by regressing the forecast errors on a constant:

$$e_t = \alpha + \varepsilon_t \tag{3}$$

The coefficient α represents the MFE, so that equation (3) is effectively a test for a zero mean of the forecasting error. The null hypothesis, that the forecasts are unbiased, would hold if $\alpha = 0$. Later in the paper we also use two modifications of equation 3 for the absolute $|e_t|$ and squared e_t^2 forecasting errors.

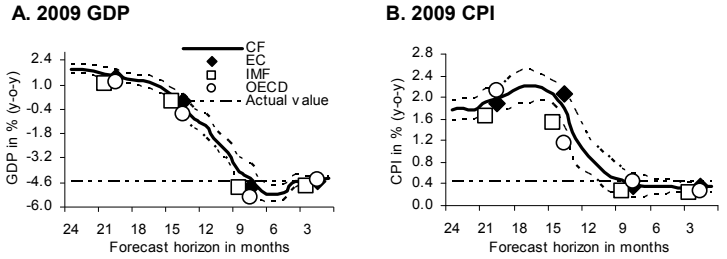
In addition, we perform a forecast comparison regression, following Ang et al. (2007) and Stock and Watson (1999):

$$a_t = \lambda f_t^{cf} + (1-\lambda) f_t^{alt} + u_t \tag{4}$$

where f_t^{cf} is the Consensus forecast, f_t^{alt} represents the alternative forecast, i.e., the international institutions' forecasts and the naive forecast, λ is the corresponding coefficient, and u_t is the forecast error associated with a combined forecast. If $\lambda = 0$, then the Consensus forecast adds nothing to the alternative forecast, and we thus conclude that the alternative forecast outperforms the Consensus benchmark. If $\lambda = 1$, then, conversely, the alternative forecast adds nothing to the Consensus forecast. It is possible that if λ is significantly negative then it does contain information but of a perverse kind. In this specific example, when the Consensus forecast is raised (lowered), the optimal combined forecast should be reduced (increased).

Finally, the Diebold-Mariano (D-M) test of statistical significance in the forecasting errors of competing forecasts is used (Diebold and Mariano, 1995). The D-M test assesses the quality of each forecast using a loss function of the forecast error. It is common to use the MSE loss differential to evaluate business cycle forecasts:

Figure 1 Effective Euro-Area GDP and CPI forecasts – visual comparison



Notes: Month 24 on the x axis represents the first available Consensus forecast for 2009. The forecast was published in the January 2008 Consensus issue. The first corresponding IMF, EC and OECD forecasts were released in April and May 2008, respectively (see the points at 21 and 20 months). Finally, the straight line represents the ex-post known actual values.

Dashed lines depict the interval containing 68 percent of the individual forecasters pooled by Consensus (one standard error around the mean Consensus forecast).

$$\bar{d} = MSE^{cf} - MSE^{alt} = 0 \tag{5}$$

where MSE^{cf} and MSE^{alt} are the mean squared errors of the Consensus and alternative forecasts, respectively. The null hypothesis (H_0) that two forecasts are on average the same is tested.

The D-M test statistic is expressed in the following form:

$$DM = \bar{d} \Big/ \left[2\pi \hat{g}_d(0) / T \right]^{1/2} \tag{6}$$

where $\bar{d}$ is the average difference between the errors of two forecasts at time $t = 1, 2, 3, \ldots T$ and $\hat{g}_d(0)$ is a consistent estimation of a distribution function $g_d(0)$.

3. Empirical Results for Effective GDP Growth and CPI and PPI Inflation

3.1 Visual Forecast Assessment

Firstly, we provide a visual forecast assessment of the last year of our sample. We find the initial visual inspection helpful before we proceed to the empirical results.

Figure 1 shows the evolution of the effective euro-area GDP growth and CPI inflation forecasts for 2009, when the financial crisis peaked. The forecasts for 2009 are characterized by the largest downward re-estimations, which are obvious given the impact of the crisis. In the case of the GDP growth forecast (see *Figure 1A*), the difference between the first published Consensus forecast (1.9 percent) and the last published Consensus forecast (-4.3 percent) reached 6.2 percentage points. The international institutions were slightly more pessimistic than Consensus.

The re-estimations were much smaller in the case of the CPI inflation forecast for 2009 (1.4 percentage points; see *Figure 1B*). The CPI inflation forecast was lifted initially due to increasing oil prices until July 2008, but afterwards it was lowered sharply as in the case of the GDP growth forecast (due to the consequences of the financial crisis).

The intervals of individual Consensus forecasters are illustrated by the dashed lines. It is apparent that a gradual shortening of the horizon implies a narrowing of

Figure 2 Gradual Improvement of Consensus Forecasts (effective euro-area GDP growth and CPI inflation, 1994–2009)

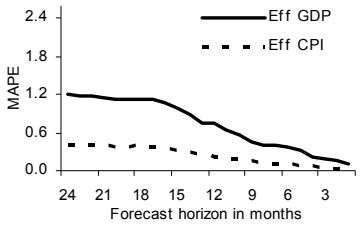
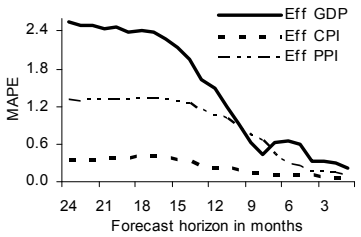


Figure 3 Gradual Improvement of Consensus Forecasts (effective euro-area GDP growth, CPI, and PPI inflation, 2004–2009)



the intervals. In other words, the individual Consensus forecasts draw nearer to each other over time, which we interpret as meaning decreasing uncertainty of the mean Consensus forecast.

Nevertheless, exceptions occur in turbulent times, when the uncertainty of the mean Consensus forecast might increase despite a shortening of the forecast horizon (for a detailed analysis see also Dovern et al., 2009).

In the case of both the GDP growth forecast and the CPI inflation forecast, the intervals around the mean Consensus forecast (the dashed lines) widened markedly after October 2008 (15 months before the end of the forecast horizon). This was associated with the increased uncertainty of the mean Consensus forecast at that time (increased disagreement among forecasters). A follow-up narrowing of the interval occurred in January 2009 in the case of CPI inflation (12 months before the end of the forecast horizon) and in June 2009 in the case of GDP growth (7 months before the end of the forecast horizon).

Subsequently, *Figures 2* and *3* show a gradual improvement in the Consensus accuracy along the forecast horizon. The mean average percentage error (MAPE) serves as a measure of the forecasting accuracy, allowing us to compare the forecasts of different economic variables. Intuitively, the forecasting errors decrease as we move toward the end of the predicted year. Forecasters have more information to make more accurate assumptions about the future and the forecasting process thus becomes easier. Similarly, Osterloh (2008) shows that the average RMSE of the Consensus forecasts for German GDP growth (across all forecasters and target years) diminishes strongly as we get closer to the end of the predicted year.

It is also apparent that Consensus is characterized by higher forecasting errors in the case of GDP growth than in the case of CPI inflation. The accuracy of the CPI

Table 1 Comparison of Forecasts: MFE

MFE	1994–2009				2002–2009			
	Effective CPI		Effective GDP		Effective CPI		Effective GDP	
	t	*t*+1	*t*	*t*+1	*t*	*t*+1	*t*	*t*+1
CF (April, October)	**0.02**	-0.12	-0.07	**-0.89*** **	**0.08**	**0.06**	-0.10	**-1.11***
IMF	0.04	**0.06**	**-0.02**	-1.03***	0.09	0.21	**0.02**	-1.17**
CF (May, November)	**0.01**	**-0.10**	-0.07	-0.85***	0.07	0.07	-0.10	-1.04*
EC	-	-	-	-	**0.00**	**0.03**	-0.06	-1.09**
OECD	-	-	**-0.03**	-0.88***	0.07	0.22	**0.00**	-1.07**
NF	-0.22	-0.26	-0.19	**-0.32**	-0.25	-0.16	-0.67	**-0.93**

Notes: - *t* is the forecast for the current year and *t*+1 is the forecast for the next year.

- MFE (mean forecast error) indicates whether a forecast is systematically biased. A positive value indicates that forecasts are, on average, underestimated.
- Equation (3) is estimated by OLS. Symbols ***, ** and * indicate rejection of the null hypothesis that the MFE is equal to zero at the 1%, 5% and 10% significance levels, respectively.

inflation forecast is very similar in both periods (compare *Figures 2* and *3*). In contrast, the accuracy of the GDP growth forecast deteriorates substantially in the shorter period from 2004 to 2009. This is caused by the outlying year 2009 and confirms the stylized fact that some years are harder to predict than others.

3.2 Empirical Results

We tested for forecast bias in the Consensus forecasts, the international institutions' forecasts, and the naive forecasts (NF). *Table 1* shows the results for the mean forecast error (MFE). The closer the value to zero, the less biased the forecast. Values in bold indicate the lowest forecasting error of the variables shown (CPI inflation and GDP growth) at the given forecast horizon (the current year and the next year).

Because the international institutions' forecasts are released twice a year only, the IMF forecasts are compared with the corresponding April and October Consensus issues, and the EC, OECD, and naive forecasts are compared with the May and November Consensus issues (see also Section 2 for clarification).

The forecasts for GDP growth are biased upward (overestimated). This bias is statistically significant for the next-year forecasts with the exception of the naive forecasts, which are characterized by the lowest mean forecasting errors. This is in accordance with the results found in the literature (Ager et al., 2009; Batchelor, 2001; Osterloh, 2008) and reflects the systematically positive expectations of forecasters about the future. Most Consensus forecasters are affiliated with investment banks and may intend to promote positive expectations among their clients (self-fulfilling expectations).

On the contrary, the CPI inflation forecasts are mostly underestimated by all institutions, including Consensus, but not at statistically significant levels, i.e., they are not biased. This holds especially for the shorter period from 2002 to 2009. This may possibly reflect a positive role of independent central banks and the resulting anchoring of inflation at low levels (Dovern et al., 2009).

It is evident from *Table 2* that all the compared forecasts are biased at the 1% significance level if we take into account the absolute forecasting error (MAFE).

Table 2 Comparison of Forecasts: MAFE

MAFE	1994–2009				2002–2009			
	Effective CPI		Effective GDP		Effective CPI		Effective GDP	
	t	*t*+1	*t*	*t*+1	*t*	*t*+1	*t*	*t*+1
CF (April, October)	**0.18*** **	0.61***	**0.4*** **	**1.29*** **	**0.18*** **	**0.54*** **	**0.34*** **	1.69***
IMF	0.2***	0.61***	0.42***	1.36***	0.21***	0.55***	0.38***	1.69***
CF (May, November)	**0.14*** **	**0.58*** **	**0.34*** **	1.22***	**0.15*** **	0.51***	**0.29*** **	1.6***
EC	-	-	-	-	0.17***	**0.5*** **	**0.29*** **	1.64***
OECD	-	-	0.36***	**1.19*** **	0.2***	0.62***	0.34***	**1.47*** **
NF	0.66***	0.92***	1.42***	1.67***	0.68***	0.70***	1.47***	2.09***

Notes: - *t* is the forecast for the current year and *t*+1 is the forecast for the next year.
- MAFE (mean absolute forecast error) indicates the average size of the forecast errors (deviations) in the examined period irrespective of the direction (positive or negative) of the error.
- Equation (3) is estimated by OLS. Symbols ***, ** and * indicate rejection of the null hypothesis that the MAFE is equal to zero at the 1%, 5% and 10% significance levels, respectively.

Table 3 Comparison of Forecasts: RMSE

RMSE	1994–2009				2002–2009			
	Effective CPI		Effective GDP		Effective CPI		Effective GDP	
	t	*t*+1	*t*	*t*+1	*t*	*t*+1	*t*	*t*+1
CF (April, October)	**0.24*** **	0.76***	**0.58** **	1.84**	**0.23** **	0.75**	**0.44*** **	2.32**
MF	0.27***	**0.74*** **	0.59***	1.84***	0.25**	**0.7*** **	0.48**	**2.22** **
CF (May, November)	**0.2*** **	**0.72*** **	0.53**	1.75**	**0.21** **	**0.7** **	0.4**	2.2*
EC	-	-	-	-	0.22**	**0.7** **	**0.39** **	2.24*
OECD	-	-	**0.50*** **	**1.66** **	0.26**	0.77**	0.44**	**2.05***
NF	0.86***	1.08***	2.03***	2.34**	0.95*	0.87***	2.19*	2.84*

Notes: - *t* is the forecast for the current year and *t*+1 is the forecast for the next year.
- RMSE (root mean squared error) penalizes larger forecast errors more.
- Equation (3) is estimated by OLS. Symbols ***, ** and * indicate rejection of the null hypothesis that the MSE is equal to zero at the 1%, 5% and 10% significance levels, respectively.

Nevertheless, regarding solely the size of the forecasting errors, the Consensus forecasts are superior to the other forecasts in most cases. Finally, the naive forecasts (NF) are characterized by having the largest forecasting errors, in accordance with our intuition.

In *Table 3* we report RMSE statistics, which are expressed in annual percentage terms. The RMSE ranges from around 0.2 percentage points in the case of the Consensus CPI inflation forecast to 2.8 percentage points in the case of the naive GDP growth forecast. As with the MAFE, all the forecasts are biased at statistically significant levels, but, again, the Consensus forecasting errors are relatively lower in many cases and, conversely, the naive forecasts are the worst.

The Consensus forecasts beat the international institutions' forecasts mainly in the current-year forecasts and, additionally, Consensus is superior to the naive forecasts in all cases. Furthermore, if we test for the null hypothesis as to whether the differences between the forecast accuracy of two competitive forecasts are on average

Table 4 Relative RMSE and D-M Test of Statistical Significance in Forecast Differences

CF vs.	1994–2009				2002–2009			
	Effective CPI		Effective GDP		Effective CPI		Effective GDP	
	t	*t*+1	*t*	*t*+1	*t*	*t*+1	*t*	*t*+1
EC	-	-	-	-	0.95	1.0	1.03	0.98***
IMF	0.89	1.03	0.98	1.0	0.92	1.07	0.92	1.05
OECD	-	-	1.06	1.05	0.81	0.91	0.91	1.07
NF	0.23***	0.67***	0.26**	0.75	0.22	0.8***	0.18	0.77

Notes :- *t* is the forecast for the current year and *t*+1 is the forecast for the next year.

- A relative RMSE value lower than 1 indicates that Consensus performs better than the alternatives (EC, IMF, OECD, and NF), whereas a value higher than 1 indicates that the alternatives are better. Stars indicate if the null hypothesis of the same forecasting accuracy of the compared forecasts can be rejected at these levels of significance: ***1%, **5%, *10%.
- The IMF forecasts are compared with the April and October Consensus issues. The EC, OECD and NF are compared with the May and November Consensus issues.

the same, taking into account the number of observations and the volatility of variables (the D-M test; see Section 2), we see that the effective CPI inflation Consensus forecasts beat the naive forecasts at the 1% significance level for both the current year and the next year (1994–2009). In the case of the GDP growth forecasts, Consensus beats the naive forecast at the 5% significance level, but for the current-year forecast only. This finding is in accordance with the previous literature (Osterloh, 2008), which concludes that the accuracy of the Consensus next-year GDP growth forecasts is relatively low.

We are not able to make any strong conclusion about the differences between the Consensus forecasts on the one hand and the IMF and OECD forecasts on the other since they are not statistically significant. Assuming only the range of the RMSE forecasting errors, Consensus is more precise than the IMF forecasts for the current year but is less accurate than the OECD forecasts for GDP growth.

Finally, all the forecasts are covered in the shorter sample (2002–2009). Consensus beats the naive forecast at the 1% significance level only in the case of the CPI inflation forecast for the next year. Nevertheless, Consensus is also superior (at the 1% significance level) to the EC forecast for GDP growth in the next year. In all remaining cases, even though the values of mostly lower than 1 in *Table 4* point to Consensus having lower forecasting errors, this finding is not statistically significant.

If we focus solely on the pre-crisis period from 1994 to 2007 (see the *Appendix, Table 4*), the results are more diverse. Most importantly, Consensus not only beats, at a statistically significant level, the EC GDP growth forecasts, but also beats the IMF GDP growth forecast for both the current and the next year. Conversely, the Consensus forecast for GDP growth for the next year is outperformed (at the 10% significance level) by the OECD forecast.

Additionally, we show the results of the information content test (*Table 5*). We are, *de facto,* carrying out a forecast comparison regression (see Equation 4, Section 2). If the presented coefficient is 1 or higher, then Consensus outperforms the alternative forecast (EC, IMF, OECD, and NF), which, in this particular case, adds nothing to a combined forecast.

Table 5 Test of the Information Content of Consensus Forecasts

CF vs.	1994–2009				2002–2009			
	Effective CPI		Effective GDP		Effective CPI		Effective GDP	
	t	*t*+1	*t*	*t*+1	*t*	*t*+1	*t*	*t*+1
EC	-	-	-	-	0.75	0.67	0.47	6.8
					(0.52)	(1.24)	(0.7)	(4.24)
IMF	0.95**	1.01***	0.58	-0.2	0.77	-0.03	0.82*	-2.9
	(0.35)	(0.36)	(0.42)	(2.08)	(0.46)	(0.7)	(0.4)	(2.09)
OECD	-	-	0.37	-2.05*	0.95**	0.79	0.67**	-3.93**
			(0.29)	(1.19)	(0.33)	(0.56)	(0.27)	(1.7)
NF	1.04***	0.71***	1.09***	1.06***	1.0***	1.33**	1.04***	1.5***
	(0.04)	(0.16)	(0.05)	(0.22)	(0.06)	(0.46)	(0.05)	(0.43)

Notes:- *t* is the forecast for the current year and *t*+1 is the forecast for the next year.

- Equation (4) is estimated by OLS. If the presented coefficient is 1 or higher, then the alternative forecasts (EC, IMF, OECD, NF) add nothing to a combined forecast of Consensus and one of the alternative forecasts. The lower the coefficient, the less the amount of information contained in the Consensus forecasts. Standard errors are given in parenthesis below the coefficient estimates. Stars indicate rejection of the null hypothesis at these levels of significance: ***1%, **5%, *10%.
- The IMF forecasts are compared with the April and October Consensus issues. The EC, OECD and NF are compared with the May and November Consensus issues.

The test reveals that Consensus outperforms the naive forecasts at the 1% significance level in all cases. In the entire sample period (1994–2009), Consensus is also superior (at statistically significant levels) to the IMF forecasts for CPI inflation in both the current and the next year. Nevertheless, it is less useful than the OECD forecast for GDP growth in the next year at the 10% significance level.

In the shorter period (2002–2009), Consensus adds more information to a combined forecast with the IMF for GDP growth in the current year at the 10% significance level and in a combined forecast with the OECD for CPI inflation and GDP growth in the current year at the 5% significance level. Conversely, Consensus adds no information, or adds information of a perverse kind, to a combined forecast with the OECD for GDP growth in the next year at the 5% significance level.

Because the international institutions (EC, IMF, and OECD) do not provide forecasts for PPI inflation, this is assessed only against the naive forecast. *Figures 4* and *5* show the forecast accuracy (measured by the mean percentage forecast error) of the April and October Consensus issues in the period from 2004 to 2009.[2] The presented variables include effective euro-area CPI and PPI inflation and GDP growth.

The Consensus April and October forecasts for the current year are characterized by lower forecasting errors (MAPE) in comparison with the naive forecasts (see *Figure 4*). The Consensus effective euro-area PPI inflation forecast for the current year is superior to the naive forecast at the 10% significance level (measured by the D-M test statistic). The forecasting errors of the next-year forecasts (*Figure 5*) match the naive forecasts except for GDP growth, where Consensus has a lower forecasting error.

[2] The short sample period is due to the limited availability of previous PPI forecasts.

Figure 4 MAPE of Current-Year Forecasts (*t*) Consensus versus NF (2004–2009)

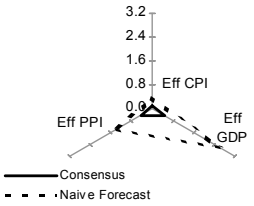
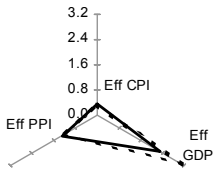


Figure 5 MAPE of Next-Year Forecasts (*t*+1) Consensus versus NF (2004–2009)



If we compare the magnitude of the forecasting errors between different variables, we find that the Consensus forecasts for GDP growth and PPI inflation for the current year are more than four times higher compared to the CPI inflation forecast. The next-year forecasting errors of GDP growth and PPI inflation are more than six and three times higher, respectively, than the CPI inflation forecasts. This reflects the different historical variability of the two classes of economic variables.

If we observe the pre-crisis period only (see the *Appendix, Figures 1 and 2*), the differences in forecasting accuracy among individual variables are not as large. Furthermore, regarding only the size of the forecasting errors, the forecast for PPI inflation for the next year outperforms the naive forecast.

3.3 Combined Forecast Accuracy

In this section we implement the method proposed by Eisenbeis, Waggoner, and Zha (2002). This method compares forecasts which contain multiple variables. In our case, we have complete forecasts for two variables: effective euro-area CPI inflation and effective euro-area GDP growth. The method is based on the fact that some variables are hard to forecast due to their high historical volatility (e.g., GDP growth), whereas other variables, like CPI inflation, which has been well anchored in recent decades, are more easily predicted. When comparing different forecasts which contain outlooks for multiple variables, it may be useful to capture their historical volatility before carrying out a joint forecast assessment of them (see *Table 6*).

In our case, the volatility (dispersion) of the actual values of effective euro--area GDP growth in the period from 1994 to 2009 was more than six times higher[3] than the volatility of effective euro-area CPI inflation. Obviously, due to its lower volatility, the forecasts of effective CPI inflation are expected to have lower forecasting errors. This is why we attach a six-times higher weighting to the forecasting errors of CPI inflation forecasts than to the GDP growth forecast. In fact, we penalize the CPI inflation forecast. After the weighting procedure we obtain a single indicator of both CPI and GDP forecasts.

Similarly to the analysis of sole variables, neither forecast significantly outperforms Consensus. Conversely, Consensus outperforms the EC forecast for the current year at the 5% significance level. Moreover, Consensus beats the naive forecast at the 1% significance level in two cases (see *Table 7*).

[3] The high difference in volatility observed between the two variables is caused by the outlying year 2009. Until 2008, the historical volatility of GDP growth was only two times higher than the volatility of CPI inflation.

Table 6 Comparison of Forecasts: Weighted Indicator (CPI, GDP)

RMSE	1994–2009		2002–2009	
	t	*t*+1	*t*	*t*+1
CF (April, October)	**0.23*****	0.79***	**0.22*****	0.86*
IMF	0.25***	**0.73*****	0.23**	**0.75****
CF (May, November)	**0.2*****	**0.75*****	**0.2****	**0.8***
EC	-	-	0.22***	0.82*
OECD	-	-	0.24**	**0.8***
NF	0.88**	1.04***	1.06*	1.05**

Notes: - *t* is the forecast for the current year and *t*+1 is the forecast for the next year.

- Weighted Indicator = $w_1.CPI + w_2.GDP$. Weights w_1 and w_2 reflect historical volatility (statistical dispersion) in the period from 1994 to 2009 ($w_1 = 0.87$, $w_2 = 0.13$).
- Equation (3) is estimated by OLS. Symbols ***, ** and * indicate rejection of the null hypothesis that the MSE is equal to zero at the 1%, 5% and 10% significance levels, respectively.

Table 7 Relative RMSE and D-M Test of Statistical Significance in Forecast Differences: Weighted Indicator (CPI, GDP)

CF vs.	1994–2009		2002–2009	
	t	*t*+1	*t*	*t*+1
EC	-	-	0.91**	0.98
IMF	0.92	1.08	0.96	1.15
OECD	-	-	0.83	1.0
NF	0.23**	0.72	0.19	0.76***

Notes:- *t* is the forecast for the current year and *t*+1 is the forecast for the next year.

- A value of relative RMSE lower than 1 indicates that Consensus performs better than the alternatives (EC, IMF, OECD, and NF), whereas a value higher than 1 indicates that the alternatives are better. Stars indicate if the null hypothesis of the same forecast accuracy of the compared forecasts can be rejected at these levels of significance: ***1%, **5%, *10%.
- The IMF forecasts are compared with the April and October Consensus issues. The EC, OECD and NF are compared with the May and November Consensus issues.

If we exclude the crisis years of 2008 and 2009, the difference between the Consensus and international institutions' forecasts is less obvious (see the *Appendix, Table 7*). None of the alternative forecasts is significantly better or worse than Consensus. Nevertheless, Consensus is superior to the naive forecast at the 1% significance level in all four cases.

Additionally, if we test for the information content of the alternative forecasts (*Table 8*), we find that Consensus is more informative than the naive forecasts at the 1% significance level. Consensus adds more to a combined forecast in the case of the IMF (1994–2009), EC, and OECD (2002–2009) current-year forecasts as well. Excluding the crisis years changes the overall picture only marginally.

Another important characteristic of a multiple-economic-variable forecast is the expected consistency among single economic variables given by their long-term relationship. We measure the mutual consistency between the CPI inflation forecast and the GDP growth forecast by their mutual correlation. The correlation coefficient between the actual values of CPI inflation and GDP growth in the period from 1994 to 2009 was positive (correlation coefficient equal to 0.51). Nevertheless, the correla-

Table 8 Test of Information Content of Consensus Forecasts: Weighted Indicator (CPI, GDP)

CF vs.	1994–2009		2002–2009	
	t	*t*+1	*t*	*t*+1
EC	-	-	1.05*	1.23
			(0.55)	(1.58)
IMF	0.83**	-0.38	0.67	-1.26
	(0.35)	(0.6)	(0.45)	(0.79)
OECD	-	-	0.95**	0.47
			(0.33)	(0.71)
NF	1.02***	1.58***	0.99***	2.06***
	(0.04)	(0.26)	(0.05)	(0.47)

Notes: - *t* is the forecast is the forecast for the current year and *t*+1 is the forecast for the next year.

Equation (4) is estimated by OLS. If the presented coefficient is 1 or higher, then the alternative forecasts (EC, IMF, OECD, NF) add nothing to a combined forecast of Consensus and one of the alternative forecasts. The lower the coefficient, the less the amount of information contained in the Consensus forecasts. Standard errors are given in parenthesis below the coefficient estimates. Stars indicate rejection of the null hypothesis at these levels of significance: ***1%, **5%, *10%.

- The IMF forecasts are compared with the April and October Consensus issues. The EC, OECD and NF are compared with the May and November Consensus issues.

tion coefficient was strongly affected by the last observation. If we exclude 2009, no correlation between these two variables is apparent.

Accordingly, the Consensus and international institutions' forecasts for the current year are correlated at similar or higher levels to the historical values. They are therefore strongly affected by the actual values. Conversely, the forecasts for the next year are less correlated. Thus, we do not detect strong consistency in the longer-term forecasts.

Furthermore, if we look at the direction of the forecast re-estimations we cannot find strong synchronization between the forecast updates of CPI inflation and GDP growth. The forecasts for these two variables are very often re-estimated in different directions. Only about 54 percent of all the Consensus re-estimations are synchronized, i.e., the forecasts for CPI inflation and GDP growth are re-estimated in the same direction. The degree of synchronization in the case of the EC, IMF, and OECD forecasts varied between 39 and 56 percent, depending on the number of observations.

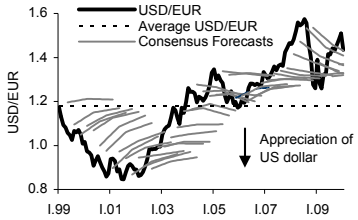
4. USD/EUR Forecast Accuracy

Figure 6 shows the evolution of the nominal USD/EUR exchange rate and the corresponding Consensus forecasts.[4] The Consensus point forecasts at the three--month, one-year, and two-year horizons are interpolated. We start with the January 1999 Consensus forecast, which extends from April 1999 to January 2001. All subsequent Consensus forecasts are constructed in the same way.

At the beginning of the observed period, when the US dollar was appreciating, the Consensus forecasts were systematically higher, i.e., a weaker dollar was sys-

[4] Only the January, April, July, and October Consensus forecasts (out of a total of twelve monthly forecasts) are presented to avoid overloading the chart.

Figure 6 USD/EUR and Its Consensus Forecasts (1999–2009)



Source: Thomson's Datastream and Consensus Economics

Table 9 Comparison of Forecasts: MFE, MAFE, RMSE

		1999–2009[a]			2002–2009[a]		
Forecast horizon		***3M***	***1Y***	***2Y***	***3M***	***1Y***	***2Y***
MFE	CF	**-0.002**	**0.006**	**0.051*****	**0.019****	**0.063*****	0.118***
	FWD	0.008	0.032***	0.077***	0.022***	0.068***	**0.114*****
	NF	0.008	0.036***	0.095***	0.021**	0.069***	0.125***
MAFE	CF	0.066***	**0.116*****	**0.146*****	0.066***	**0.101*****	**0.125*****
	FWD	**0.06*****	0.119***	0.152***	**0.063*****	0.123***	0.135***
	NF	0.062***	0.118***	0.153***	0.068***	0.127***	0.145***
RMSE	CF	0.081***	0.138***	**0.177*****	0.081***	**0.122*****	**0.157*****
	FWD	**0.075*****	0.137***	0.188***	**0.079*****	0.141***	0.177***
	NF	0.079***	**0.135*****	0.187***	0.085***	0.144***	0.186***

Notes: - [a] *3M* (3 months) ahead forecasts are assessed until the December 2009 forecast, *1Y* (1 year) ahead forecasts are assessed until the March 2009 forecast and *2Y* (2 years) ahead forecasts are assessed until the March 2008 forecast.

- *CF:* Consensus forecast, *FWD:* forecast derived from forward exchange rates on the survey day of the Consensus forecast, *NF*: naive forecast.
- Equation (3) is estimated by OLS. Symbols ***, ** and * indicate rejection of the null hypothesis that the MSE is equal to zero at the 1%, 5% and 10% significance levels, respectively.

tematically predicted. As from 2002, when the trend reversed, the dollar was depreciating faster than Consensus assumed. When the dollar moved above its average value for the whole period (USD/EUR 1.18), Consensus started to forecast stability or appreciation of the dollar. Finally, after the dollar reached historical lows (in summer 2008) Consensus started to systematically predict its appreciation.

Restricted to our sample period from 1999 to 2009, the Consensus forecasts tended systematically toward the average USD/EUR value. Accordingly, the disparity between the one-year and two-years-ahead forecasts was negligible. They both reached USD/EUR 1.2 on average.

From *Table 9* (MFE), it is apparent that all forecasts were unbiased only in the case of the three-months-ahead forecasts in the long sample from 1999 to 2009. In addition, the one-year-ahead Consensus forecast was unbiased as well. The remaining forecasts were biased downward, i.e., a stronger dollar was systematically predicted.

Table 10 Relative RMSE and D-M Test of Statistical Significance in Forecast Differences

CF vs.	1999–2009[a]			2002–2009[a]		
	3M	*1Y*	*2Y*	*3M*	*1Y*	*2Y*
FWD	1.08**	1.01	0.94	1.03	0.87**	0.89
NF	1.03	1.02	0.95	0.95*	0.85***	0.84*

Notes: - [a] *3M* (3 months) ahead forecasts are assessed until the December 2009 forecast, *1Y* (1 year) ahead forecasts are assessed until the March 2009 forecast and *2Y* (2 years) ahead forecasts are assessed until the March 2008 forecast.

- *CF*: Consensus forecast, *FWD*: forecast derived from forward exchange rates on the survey day of the Consensus forecast, *NF*: naive forecast.

- A value of relative RMSE lower than 1 indicates that Consensus performs better than the alternatives (EC, IMF, OECD, and NF), whereas a value higher than 1 indicates that the alternatives are better. Stars indicate if the null hypothesis of the same forecast accuracy of the compared forecasts can be rejected at these levels of significance: ***1%, **5%, *10%.

Table 11 Test of Information Content of Consensus Forecasts

CF vs.	1999–2009[a]			2002–2009[a]		
	3M	*1Y*	*2Y*	*3M*	*1Y*	*2Y*
FWD	-0.22	0.44*	0.94***	0.19	1.01***	2.04***
	(0.23)	(0.22)	(0.25)	(0.37)	(0.06)	(0.37)
NF	0.22	0.41**	0.79***	1.44***	1.87***	2.45***
	(0.28)	(0.19)	(0.2)	(0.43)	(0.27)	(0.31)

Notes: Equation (4) is estimated by OLS. If the presented coefficient is 1 or higher, then the alternative forecasts (FWD and NF) add nothing to a combined forecast of Consensus and one of the alternative forecasts. The lower the coefficient, the less the amount of information contained in the Consensus forecasts. Standard errors are given in parenthesis below the coefficient estimates. Stars indicate rejection of the null hypothesis at these levels of significance: ***1%, **5%, *10%.

The Consensus forecasting accuracy was compared with the naive forecast and the forecast derived from the forward exchange rate (FWD). Both forecasts were even more biased, except for the FWD forecast at the two-year horizon in the short sample period (2002–2009).

Looking at *Table 10*, the Consensus three-months-ahead forecast is outperformed by the FWD forecast at the 5% significance level in the whole sample (1999––2009). Nevertheless, in the short sample period from 2002 to 2009, the Consensus forecast improved relative to both the FWD and the naive forecast. Consensus was, in particular, superior to the naive forecast at all three forecast horizons and, in addition, against the FWD forecast at the one-year horizon.

If we exclude the crisis years (see the *Appendix, Table 7*), the results are the same in the long sample period from 1999 to 2007. Nevertheless, the results for the short sample period from 2002 to 2007 confirm the statistically significant superiority of the Consensus forecast (in accordance with *Table 10*).

Furthermore, the results of the information content test (*Table 11*) support the superiority of Consensus. In the short sample period Consensus is superior to the alternative forecasts at the 1% significance level at all forecast horizons except for the FWD three-months-ahead forecast. In the long sample period, Consensus

outperforms the alternative forecasts in the two-years-ahead forecast at the 1% significance level.

5. Conclusion

We compared the accuracy of the Consensus Economics forecasts to those of the European Commission, International Monetary Fund, and Organization for Economic Co-operation and Development, and also to the naive forecast and the forecast implied by forward exchange rates. Basic descriptive statistics of forecast efficiency and forecast bias as well as tests for statistical significance in the differences between competing forecasts were applied. We analyzed the forecasts for GDP growth, CPI inflation, PPI inflation, and USD/EUR.

Forecasting accuracy was compared before and during the recent financial crisis. With respect to the GDP growth and CPI inflation forecasts, we found that the Consensus Economics forecasts usually outperformed the others significantly, especially during the whole period from 1994 to 2009. The results were not that strong for the pre-crisis period. The Consensus forecasts beat the international institutions' forecasts mainly in the current-year forecasts. In the next-year forecasts, the IMF and OECD forecasts perform similarly to those of Consensus. Additionally, Consensus is superior to the naive forecasts at all forecast horizons.

In accordance with previous literature, we confirm a relatively low level of accuracy of the next-year GDP growth forecasts, which are biased upward by all institutions, whereas the CPI inflation forecasts are unbiased.

The PPI inflation and USD/EUR forecasts were compared only with the naive forecast and the forward rate forecast in the case of USD/EUR due to the absence of comparable forecasts by international institutions. The Consensus PPI inflation forecast significantly beats the naive forecast for the current year. The Consensus forecast for USD/EUR improved dramatically after 2002, when it significantly outperformed the naive forecast at all forecasting horizons and the forward implied forecast at the one-year horizon. A stronger dollar was systematically predicted over the observed period by all the assessed forecasts.

The practical advantage of the Consensus Economics forecasts lies in a broader range of predicted economic variables and in the high frequency of forecast releases (every month). Based on the relatively high forecasting accuracy and the additional practical characteristics of Consensus, the Czech National Bank will continue to use it in its prediction process. However, the Consensus Forecasts will also be regularly confronted with alternative assumptions.

REFERENCES

Ager P, Kappler M, Osterloh S (2009): The Accuracy and Efficiency of the Consensus Forecasts: A Further Application and Extension of the Pooled Approach. *International Journal of Forecasting,* 25(1):167–181.

Ang A, Bekaert G, Wei M (2007): Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better? *Journal of Monetary Economics*, 54:1163–1212.

Antal J, Hlaváček M, Horváth R (2008): Do Central Bank Forecast Errors Contribute to the Missing of Inflation Targets? The Case of the Czech Republic. *Finance a úvěr-Czech Journal of Economics and Finance*, 58(9–10):434–453.

Babecký J, Podpiera J (2011): Inflation Forecast Errors in the Czech Republic: Evidence from a Panel of Institutions. *Eastern European Economics*, 49(1):72–83.

Batchelor R (2001): How Useful are the Forecasts of Intergovernmental Agencies? The IMF and OECD Versus the Consensus. *Applied Economics*, 33(2):225–235.

Batchelor R (2007): Bias in Macroeconomic Forecasts. *International Journal of Forecasting*, 23(2):189–203.

Bowles C, Fritz R, Genre V, Kenny G, Meyler A, Rautanen T (2007): The ECB Survey of Professional Forecasters – A Review after Eight Years' Experience. *ECB Occasional Paper Serie*, no. 59.

Croushore D (2010): An Evaluation of Inflation Forecasts from Surveys Using Real-Time Data. The *B.E. Journal of Macroeconomics*, 10(1):1–30.

Diebold FX, Mariano RS (1995): Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Dovern J, Fritsche U, Slacalek J (2009): Disagreement among Forecasters in G7 Countries. *ECB Working Paper*, no. 1082.

Eisenbeis R, Waggoner D, Zha T (2002): Evaluating Wall Street Journal Forecasters: A Multivariate Approach. *Federal Reserve Bank of Atlanta*, *WP*, no. 2002–8a.

Osterloh S (2008): Accuracy and Properties of German Business Cycle Forecasts. *Applied Economics Quarterly*, 54(1):27–57.

Stock J, Watson M (1999): Forecasting Inflation. *Journal of Monetary Economics*, 44(2):293–335.

Timmermann A (2006): An Evaluation of the World Economic Outlook Forecasts. *IMF Working Paper*, no. 06/59.