ifo Working Papers

Assessing the Forecast Properties
of the CESifo World Economic Climate Indicator:
Evidence for the Euro Area

Oliver Hülsewig

Johannes Mayr

Stéphane Sorbe

Assessing the Forecast Properties of the CESifo World Economic Climate Indicator: Evidence for the Euro Area*

Abstract

This paper evaluates short-term forecasts of real GDP in the Euro area derived from the CESifo Economic Climate indicator (WES) in terms of forecast accuracy. We compare the forecast properties of the WES with those of monthly composite indicators. Considering the WES is interesting because (i) it is exclusively based on the assessment of economic experts about the current economic situation, and (ii) it is timely released within the quarter on a quarterly basis. The empirical analysis is carried out under full information, which means that the competing monthly indicators are known for the entire quarter, and under incomplete information. Our findings exhibit that the forecast power of the WES is comparatively proper.

Oliver Hülsewig
Ifo Institute for Economic Research
at the University of Munich
Poschingerstr. 5
81679 Munich, Germany
Phone: +49(0)89/9224-1689
huelsewig@ifo.de

Johannes Mayr
Ifo Institute for Economic Research
at the University of Munich
Poschingerstr. 5
81679 Munich, Germany
Phone: +49(0)89/9224-1228
mayr@ifo.de

Stéphane Sorbe
Institut National de la Statistique
et des Etudes Economiques
(INSEE), Paris.
France
stephane.sorbe@polytechnique.org

1 Introduction

Obtaining short–term projections of real GDP from business–cycle indicators guarantees that timely information is explicitly exploited. These indicators include quantitative indicators, such as industrial production, confidence surveys and composite indicators. The forecast properties of business–cycle indicators have been examined by Parigi and Schlitzer (1995), Camba–Mendez et al. (2001), Baffigi, Golinelli, and Parigi (2002), Banerjee, Marcellino, and Masten (2003), Mourougane and Roma (2003), Rünstler and Sédillot (2003), Sédillot and Pain (2003), Gayer (2005) and Golinelli and Parigi (2007) for a number of OECD countries, which has shown that short–term forecasts of real GDP growth derived from such indicators usually perform properly.

Since Eurostat publishes the first official release of quarterly real GDP in the Euro area with a delay of several weeks, timely information about the state of the economy is appreciable. In addition to the quantitative indicators, certain composite indicators provide an insight. These include the economic sentiment indicator (ESI) of the European Commission, the OECD composite leading indicator (OLI) and the EuroCOIN indicator (ECI) by the CEPR that are calculated on a monthly basis by extracting the information contained in different quantitative indicators, confidence surveys, price indices and financial variables. Additionally, the CESifo Economic Climate indicator (WES) for the Euro area provides an assessment of economic experts about the current economic situation and their expectations.

This paper evaluates short–term forecasts of real GDP in the Euro area derived from the WES in terms of forecast accuracy. We compare the forecast properties of the WES with those of the ESI, OLI and the ECI. Focusing on the WES is interesting as it contains two specific features that are in contrast to the composite indicators: (i) it is exclusively based on the judgment of economic experts, and (ii) it is timely released within the quarter on a quarterly basis. A continuous monthly update of fresh monthly information within the survey quarter thus becomes impossible. A priori this suggests that the forecast accuracy of the WES is comparatively minor.[1]

We derive quarterly projections of real GDP from the competing indicators by estimating bridge models on the basis of a recursive regression procedure, which allows us to conduct a series of pseudo one–quarter–ahead out–of–sample forecasts. We explore the forecast properties of the indicators by means of standard forecast performance tests, which include the Root Mean Squared Forecast Error, the forecast accuracy test by Harvey, Leybourne, and Newbold (1997) – that is a

---

[1]Although, a number of studies for the Euro area have explored the forecast properties of a variety of business–cycle indicators, the WES has not yet been considered.

modified version of the Diebold and Mariano (1995) test – and a turning point test developed by Pesaran and Timmermann (1992) that allows us to judge forecast directional correctness. We select an AR–model for real GDP growth to obtain the benchmark projection.

As in Golinelli and Parigi (2007) and Rünstler and Sédillot (2003) our comparison of the forecast performance of the indicators is twofold. In the first step, we generate pseudo out–of–sample forecasts of real GDP growth under the assumption of full information, which means that the indicators are known for the entire three months within the current quarter. In the second step, we derive pseudo out–of–sample forecasts of real GDP growth by focusing on incomplete information, which implies that the monthly indicators – i.e. the ESI, OLI and the ECI – are only partially available within the current quarter. As a consequence, these indicators have to be extrapolated to generate the missing observations for the quarterly value, which exposes additional uncertainty.

Our findings suggest that the WES is an accurate forecast measure that is capable to provide a sound understanding of the actual economic situation at a relatively early moment in the quarter. The forecast properties of the WES are similar to those of the OLI, which constitutes the dominant composite indicator in terms of forecast accuracy. A comparison between the forecasts performance of the WES and Consensus Forecast on the basis of real time data provides robustness of the results by showing that the rival predictions perform equally proper.

The remainder of the paper is organized as follows. Section 2 sets out an overview of bridge models, introduces our data set for the Euro area and briefly discusses the forecast performance tests applied. In Section 3, the forecast evaluation is presented. First, we assess out–of–sample forecasts of real GDP derived from the candidate indicators (i) for the case of full information and (ii) for the case of incomplete information. The forecasts are evaluated by means of the forecast performance tests. Second, we compare the forecast properties of the WES and Consensus Forecast by using real time data. Section 4 provides concluding remarks.

2 Modeling Approach, Choice of Data and Forecast Performance Tests

2.1 Quarterly Bridge Models

Usually, bridge models are based on an Autoregressive Distributed Lag model of the form (Banerjee, Marcellino, and Masten (2003)):

$$A(L)Y_t = \delta + \sum_{j=0}^{n} B_j(L)X_{jt} + \varepsilon_t, \tag{1}$$

where Y_t denotes real GDP expressed in quarterly growth rates, δ is a constant term, X_{jt} are the quarterly values of the business–cycle indicators, $A(L)$ and $B_j(L)$ describe lag polynomials and ε_t are residuals that are assumed to be white noise. Quarterly predictions of real GDP growth are derived by exploiting the timely information contained in the indicators.

The application of bridge models to generate short–term forecasts of real GDP can be carried out either under the assumption that the indicators are completely available for the current quarter or under the assumption that the indicators are only partially known, which means that information is only disposable for the first months of the quarter. This requires the indicators to be extrapolated to obtain the missing monthly observations for the entire quarter. Three different situations can be distinguished (Golinelli and Parigi, 2007):

1. Quarterly forecasts of real GDP with indicators that are completely unknown. In this case the indicators have to be extrapolated three months into the future to derive the quarterly values.

2. Quarterly forecasts of real GDP derived from indicators that are known for the first month of the current quarter, which means that the monthly series need to be extrapolated for two months.

3. Quarterly forecasts of real GDP derived from indicators that are known for the first two months of the current quarter, which implies that the monthly series need to be extrapolated only for one month.

In the run–up of the forecast exercise the extrapolated values of the monthly series have to be aggregated to obtain the quarterly value. The aggregation scheme can be based on the mean value of the monthly data.

Obviously, obtaining quarterly projections of real GDP from indicators that are released on a monthly basis is exposed to additional uncertainty, which stems

from the necessity of extrapolating the monthly series under incomplete information. Using indicators that are published on a quarterly basis possibly avoids this ambiguity, but at the expense of less up–to–date information since a continuous monthly update becomes impossible.

2.2   Data Selection

Our data set for the Euro area comprises real GDP and various business–cycle indicators for the sample period from 1991Q1 to 2006Q3. Real GDP is seasonally adjusted and transformed into quarterly growth rates. The business–cycle indicators are grouped into quantitative and qualitative indicators:

1. The set of quantitative indicators includes industrial production (IP), new car registrations (CAR) and industrial production in construction (IPC), which are collected from Eurostat, and additionally retail sales (RS), which is taken from the OECD.[2] The data is seasonally adjusted and transformed into quarterly growth rates.

2. The qualitative indicators comprise the CESifo Economic Climate indicator (WES) for the Euro Area and three composite indicators, namely the economic sentiment indicator (ESI) of the European Commission, the OECD composite leading indicator (OLI) and the EuroCOIN indicator (ECI) of the CEPR, which are widely acknowledged and readily available. As the qualitative indicators are constructed to fluctuate around a constant mean and thus are considered to be mean stationary, their level values are implemented. Figure 1 depicts quarterly real GDP growth in conjunction with the qualitative indicators.

   The WES summarizes the assessments of economic experts on the economic situation and outlook. It is exclusively based on qualitative information and is timely published on a quarterly basis within the survey quarter.[3] The ESI combines the weighted information contained in several confidence indicators, such as industrial, service and consumer surveys (European Commission, 2007). The OLI is derived from an aggregation of a number of national indicators, which include survey data, several quantitative indicators, price indices, financial variables and

---

[2]Since Eurostat provides information on retail sales not before 1995, we decided to include OECD data.

[3]The WES is calculated as the arithmetic mean of the assessment of the economic situation in the current quarter and the expectations about the economic situation in the coming two quarters. The indicator reflects the responses of about 275 experts. See Stangl (2007) for an overview.

Figure 1: Qualitative Indicators and Quarterly Real GDP Growth

6
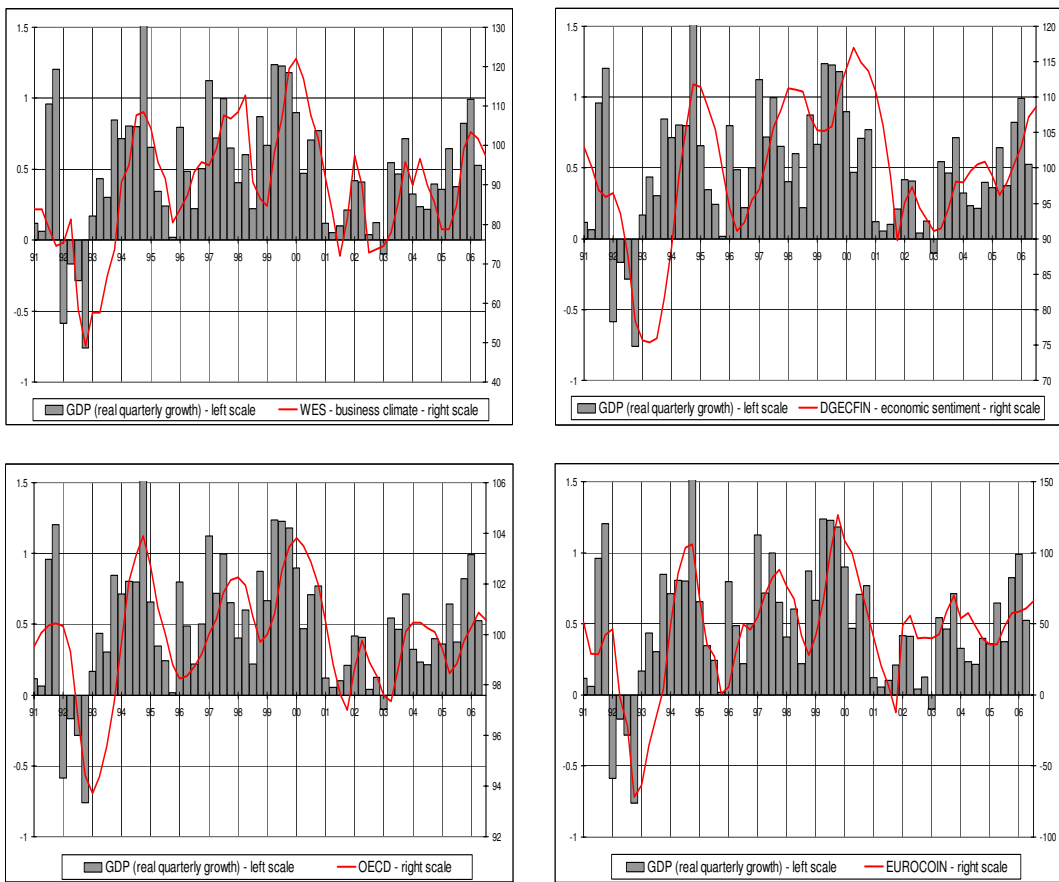
the terms of trade (OECD, 2003). Finally, the ECI is constructed from a dynamic factor analysis of an intensive number of business–cycle indicators with the purpose to track the principal common factor of aggregate economic activity (Altissima, et al., 2001). While the WES is released on a quarterly basis, the composite indicators are published monthly.

Figure 2: Stylized Overview of Relevant Events
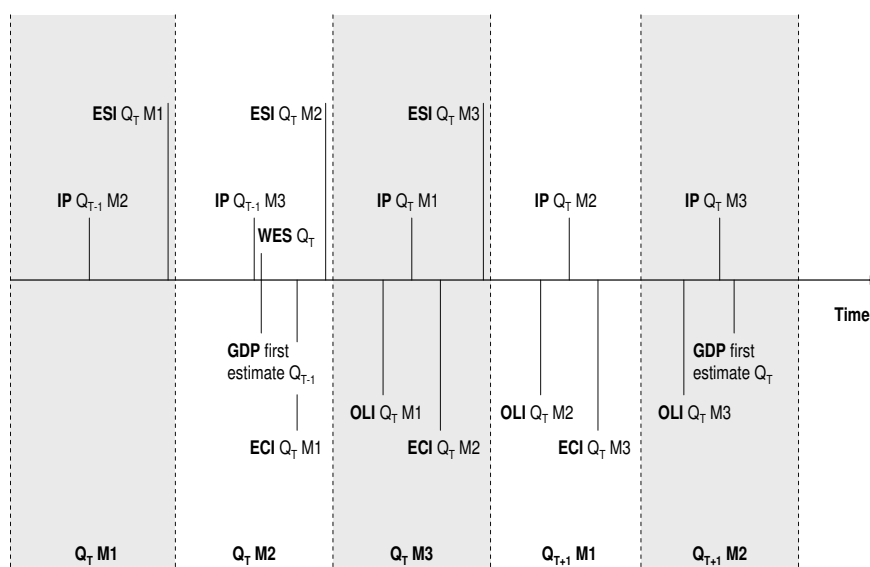
Notes: IP: industrial production; ESI: economic sentiment indicator; WES: CESifo Economic Climate indicator for the Euro area; ECI: EuroCOIN indicator; OLI: OECD leading composite indicator. Q_T denotes the current quarter; Mx denotes the respective months of the quarter ($x = 1, 2, 3$).

For the production of short–term forecasts of real GDP in real time, Figure 2 presents a stylized overview of relevant events. The first release of real GDP growth for the current quarter Q_T is published in the middle of the second month M2 of the next quarter Q_{T+1}. Usually, the set of indicators is completely available by then. IP is released with a delay of about six weeks, which implies that industrial production for Q_TM1 – as an example – is issued in Q_TM3. The WES is issued in the middle of the second month M2 of the current quarter Q_T, while the ESI is published at the end of each month, which means that the indicator for the current quarter Q_T is completely available at the end of Q_TM3. The ECI exhibits a post–carriage of two to three weeks. The OLI is released with a delay of about six weeks, which implies that the indicator for the current quarter Q_T

is completely available not until the second month M2 of the next quarter Q_{T+1}. For the creation of forecasts this timing of events has to be taken into account.

2.3 Forecast accuracy tests

We evaluate the forecast properties of the candidate indicators by means of a number of forecast performance tests that refer to forecast accuracy and forecast direction correctness. The out–of–sample Root Mean Squared Error (RMSE) is employed as a descriptive measure, which provides an indication of the accuracy of a forecast by stating that projections with a lower value are preferable. In addition, we apply the test of Harvey, Leybourne, and Newbold (HLN) (1997) that evaluates the differences of forecast errors derived from point forecasts of competing models for statistical significance.

The HNL (1997) test is a modified version of the test developed by Diebold and Mariano (1995) that is corrected for a small sample bias. The null hypothesis of equality of the expected forecast performance of two competing models is formulated as:

$$H_0 : E\left[\delta_t\right] = 0, \tag{2}$$

where the sequence of loss differentials δ_t is defined by: $\delta_t = g(e_{it}) - g(e_{jt})$. The loss functions $g(e_{it})$ and $g(e_{jt})$ are derived from the forecast errors e_{it} and e_{jt} of the rival models. Although the test allows for a wide class of prediction accuracy measures, we restrict the analysis to the out–of–sample forecast RMSE to specify the loss functions. The test is based on the following statistic:

$$\text{HLN} = \text{DM}\left[\frac{N + 1 - 2h + h(h-1)/N}{N}\right]^{\frac{1}{2}}, \tag{3}$$

where DM denotes the standard statistic of the Diebold and Mariano (1995) test, N is the number of independent point forecasts and h denotes the forecast horizon. The test compares the HLN statistic to a critical value that is drawn from a Student's t–distribution with $N - 1$ degrees of freedom.

Finally, we employ the turning point (TP) test proposed by Pesaran and Timmermann (1992) to evaluate forecast directional accuracy since obtaining information on the expected direction of movements in real GDP growth is also valuable. The TP test is a distribution–free procedure that is based on the proportion of times that the direction of change in the target variable y_t is correctly predicted by the time series of forecasted values x_t in any underlying sample. It involves a comparison to a naive coin flip as the benchmark model and only requires information on the direction of change of the target time series and the time series of forecasted values. The test is based on the standardized binomial

8

variate, which is asymptotically distributed as $N(0, 1)$. The procedure is valuable for a wide class of underlying probability distributions, as it only postulates that the probability of changes in the direction of y_t and x_t is time–invariant. We implement the test by focusing on the quarter on quarter direction of change in real GDP growth.

3 Out–Of–Sample Forecast Evaluation

We generate quarterly forecasts of real GDP from the candidate indicators by estimating the bridge models (1) recursively over the forecast sample from 2001Q1 to 2006Q3. The forecasts are derived as one–quarter–ahead out–of–sample predictions for each quarter following the starting sample from 1991Q1 to 2000Q4, that is stepwise augmented by including an additional quarter.[4] We evaluate the forecast properties of the indicators by means of the forecast performance tests, which are based on the forecast errors of 23 out–of–sample predictions. We select an AR(1)–process for real GDP growth to obtain the benchmark projection.[5]

As in Golinelli and Parigi (2007) and Rünstler and Sédillot (2003), our evaluation of the forecast performance of the indicators is two–fold. First, we explore pseudo out–of–sample forecasts of real GDP growth by focusing on full information, which implies that all indicators are known for the entire quarter. Second, we examine pseudo out–of–sample forecasts of real GDP growth by considering the moment of the release of the WES in the quarter, which means that the monthly indicators are only partially available. Since the monthly indicators need to be extrapolated, we investigate the use of various auxiliary forecast models that include a naive projection,[6] an univariate autoregressive moving average (ARMA) model, a vector autoregressive (VAR) model and a Bayesian VAR (BVAR) model, all of which are adequate to account for the staggered timing of the monthly data releases.

Our forecast exercise is based on a variety of bridge models for the candidate indicators that vary in the choice of the lag length. Following Granger (1993), we chose those specifications that provide the lowest value of the out–of–sample forecast RMSE under complete information as a criterion of model selection,

---

[4]The bridge models for each candidate indicator are estimated by including an impulse dummy. The dummy variable accounts for an outlier in quarterly real GDP growth and takes the value of one in 1995Q1 and otherwise zero.

[5]The inspection of the correlogram of quarterly real GDP growth strongly suggests the specification of an AR(1)–process. In addition, we find that the AR(1)–model unambiguously dominates competing ARIMA models in terms of the out–of–sample forecast RMSE.

[6]In the naive projection approach, the missing monthly observations are derived by means of a random walk forecast, i.e. the values depend only on the last known monthly data point.

since in–sample selection measures – such as the standard information criteria – frequently fail to provide strong implications for the out–of–sample performance.

3.1 Predictions of real GDP under Full Information

3.1.1 Indicators taken singly

Our comparison of the forecast properties of the candidate indicators starts by focusing on the case of full information. For each indicator, Table 1 displays the outcome of the forecast performance tests, which are based on the one–quarter–ahead out–of–sample forecast errors.

Table 1: Forecast Properties of the Indicators taken singly

		RMSE	HLN–Test	TP–Test	p–value
Quantitative indicators					
Industrial production	IP	0.21	−1.37	12	0.34
Retail sales	RS	0.30	+1.50	14	0.11
Car registration	CAR	0.28	+1.06	14	0.11
Ind. prod. construction	IPC	0.28	+0.84	14	0.11
Qualitative indicators					
CESifo Economic Climate	WES	0.22	−1.52	15	0.05
OECD Leading indicator	OLI	0.20	−2.24	16	0.02
Economic sentiment	ESI	0.24	−0.61	15	0.05
EuroCOIN indicator	ECI	0.26	−0.08	13	0.20
Benchmark forecast					
AR(1) model	AR	0.26	–	13	0.20

Notes: For the HLN (1997) test the corresponding critical value is ±1.31 for the 5% level with 22 degrees of freedom. A value of the HLN statistic below -1.31 implies an improvement, while a value above +1.31 implies a worsening of the forecast compared to the AR(1) benchmark prediction. TP denotes the number of correctly identified changes in the direction of real GDP growth; the p–value denotes statistical significance.

Industrial production constitutes the sole quantitative indicator that – as indicated by the HLN (1997) test – outperforms the AR(1) benchmark forecast significantly. The same applies to the WES, which equally fulfills forecast accuracy but also represents a proper measure for correctly predicting turning points. The OLI surpasses the competing composite indicators by improving upon the AR(1) benchmark prediction unambiguously. Likewise the OLI is appropriate - similar to the ESI - for accomplishing forecast directional correctness.

The forecast performance of the ECI is comparatively poor. This finding is sharply in contrast with the results of Rünstler and Sédillot (2003), who conclude that the EuroCOIN indicator constitutes the best composite indicator in terms of forecast accuracy by focusing on the forecast sample from 1998Q1 to 2001Q4. Accordingly, this suggests that the forecast power of an indicator can vary considerably over time (see also Baffigi, Gionelli and Parigi, 2004).

3.1.2 Encompassing regressions

Short–term forecasts of real GDP derived from IP under complete information are possibly enhanced by additionally accounting for the qualitative indicators.[7] We explore this conjecture by running a test of forecasting encompassing, which compares the accuracy of two rival forecasts.

Following Clements and Harvey (2006), the test is based on the regression equation:

$$y_t = \alpha f_{1t} + (1 - \alpha) f_{2t} + u_t,$$

where y_t denotes the reference series that is forecasted through a linear combination of the rival forecasts f_{1t} and f_{2t} with a combined forecast error u_t. The null hypothesis that f_{1t} is encompassed by f_{2t} is: $H_0 : \alpha = 0$, which implies that f_{2t} contains all the useful information in f_{1t}. The alternative hypothesis is typically one–sided, i.e. $\alpha > 0$. Table 2 summarizes the outcome.

Table 2: Encompassing regression against IP

		Estimated α	Std. Dev.
CESifo Economic Climate	WES	0.43	0.19
OECD Leading indicator	OLI	0.57	0.23
Economic sentiment	ESI	0.25	0.28
EuroCOIN indicator	ECI	-0.04	0.31

Notes: Test of forecasting encompassing of two rival forecasts. The null hypothesis that the forecast of a qualitative indicator is encompassed by the forecast of industrial production is rejected when α is significantly larger than zero.

The findings show that forecasts of real GDP growth generated by IP benefit form the additional information contained in the WES since the null hypothesis of forecast encompassing is clearly rejected. The same holds for the OLI, while for the ESI and the ECI the estimated parameter α is not significantly different

---

[7]Since the forecast properties of RS, CAR and IPC are relatively poor, we ignore the use of these indicators in the following.

from zero. This supports the notion that the WES and the OLI constitute the superior qualitative indicators as measured in terms of forecast accuracy.

3.1.3 Combined forecast models

Deriving forecasts of real GDP from industrial production combined with an individual qualitative indicator might give a deeper insight into the predictive power of the rival series.[8] Table 3 summarizes the results of different forecast performance tests. The HLN (1997) test compares the combined IP forecasts with the pure IP forecasts by evaluating the differences of the forecast errors for statistical significance.

Table 3: Combined Forecast Models

	RMSE Ratio	HLN–Test
IP + CESifo Economic Climate	0.95	−0.35
IP + OECD Leading indicator	0.95	−0.89
IP + Economic sentiment	1.01	−0.27
IP + EuroCOIN indicator	1.14	+1.06

Notes: RMSE of the combined IP forecast in ration to the benchmark RMSE of the pure IP forecast. For the HLN (1997) test the corresponding critical value is ± 1.31 for the 5% level with 22 degrees of freedom. A value of the HLN statistic below -1.31 implies an improvement, while a value greater that 1.31 implies a worsening of the forecast compared to the benchmark prediction.

Short-term forecasts of real GDP generated by IP combined with the WES lead to an improvement of the out–of–sample forecast RMSE that declines slightly. This also applies to the OLI, but not to the ESI and the ECI, which confirms our results of the encompassing regressions. However, the HNL (1997) test indicates that the forecasts from the combined IP models are not unambiguously superior. Since this suggests that the gains of combined models are only minor, we continue to focus on the indicators taken singly.

So far, our evaluation of the forecast performance of the indicators has built on the assumption of full information, which establishes the most convenient environment for the monthly indicators in the sense that their forecast power ought to decline when less information is available. Next, we turn to an assessment of this issue.

---

[8]This leads to various model specifications that differ in the lag structure. Again as a criterium for model selection, we chose those specifications that produce the lowest out–of–sample forecast RMSE.

3.2 Forecasting real GDP under Incomplete Information

Obtaining a first prompt forecast of real GDP from the candidate indicators at an early moment in the quarter contributes to a sound understanding of the actual economic situation. As we aim at evaluating the forecast performance of the WES, we consider the moment of the release of that indicator, which usually takes place – as shown in Figure 2 – in the middle of the second month of the quarter. As a consequence, the monthly indicators have to be extrapolated since they are almost completely unknown. Only the ESI is available for the first month of the quarter.

The necessity of forecasting the monthly indicators exposes additional uncertainty. Since the forecast performance of the monthly indicators crucially depends on the quality of the monthly predictions, we investigate the application of several auxiliary forecast models that are capable of accounting for the delayed releases of the monthly series.

Our forecast exercise under incomplete information proceeds in two steps. First, we derive forecasts of the monthly indicators from the different auxiliary forecast models. Second, we investigate the forecast performance of the indicators at the moment of the release of the WES by using the extrapolated monthly series.

3.2.1 Predicting the monthly indicators

We generate forecasts of the monthly indicators by using several auxiliary forecast models that include a naive projection, univariate ARMA models, VAR models and BVAR models.[9] Rünstler and Sédillot (2003) find that BVAR models perform well in terms of the out–of–sample forecast RMSE, closely followed by VAR models and ARMA models that also establish a firm ground as regards forecast accuracy.[10] Diron (2006) states that especially ARMA models constitute a convenient forecast device in terms of forecast exactness.

The predictions of the monthly indicators derived from the auxiliary forecast models embrace three–month–ahead forecasts for IP, the OLI and the ECI, while for the ESI two–month–ahead forecasts are established. The forecast models are specified with varying lag lengths. The VAR models include all candidate indicators to make efficient use of the entire information available.[11] The BVAR models are set up with the standard Minesota priors – as proposed by Doan,

---

[9]We use an ARIMA model for IP and ARMA models for the monthly composite indicators.

[10]Rünstler and Sédillot (2003) find that BVAR models outperform the competing auxiliary forecast models especially for longer forecast horizons of up to six months.

[11]In addition, we have considered various other business–cycle indicators, such as confidence surveys, financial variables and the terms of trade which, however, have not lead to an improvement of the forecasts.

Litterman, and Sims (1984) – which impose restrictions by assuming that the endogenous variables follow a random walk. As a criterium of model selection we chose those specifications that produce the lowest value of the out–of–sample forecast RMSE.

We forecast the monthly indicators by estimating the auxiliary forecast models recursively over the forecast sample from January 2001 to September 2006. The forecasts of the monthly indicators are derived as out–of–sample predictions for the respective months of each quarter following the starting sample from January 1991 to December 2000 that is continuously expanded by adding the next months of the subsequent quarter. We evaluate the forecasts of the monthly indicators by focusing on the out–of–sample forecast RMSE that results from the aggregate quarterly values of the forecasted monthly series.[12] Table 4 displays the outcome. For each indicator, the best auxiliary forecast model is marked by an asterisk.

Table 4: Performance of quarterly indicator forecasts

	Naive Projection	ARMA	VAR	BVAR
Industrial production[a]	1.00	1.06	0.97	0.96*
OECD indicator[a]	1.00	0.69*	0.82	0.81
Economic sentiment[b]	1.00	0.85*	0.96	0.92
EuroCOIN indicator[a]	1.00	0.93*	1.01	0.99

Notes: Measured in terms of the out–of–sample forecast RMSE relative to the naive projection. Industrial production in monthly growth rates, all other indicators in levels. The best auxiliary forecast model evaluated in terms of the lowest out–of–sample forecast RMSE is indicated by an asterisk. [a]Three step ahead forecasts. [b]Two step ahead forecasts.

Forecasts of industrial production resulting from the BVAR model predominate in terms of the out–of–sample forecast RMSE. This is in line with Rünstler and Sédillot (2003), who report a similar finding. For the composite indicators the specified ARMA models provide the lowest out–of–sample forecast RMSE, which implies that these models are preferable. Not surprisingly the naive projections come off badly. Building on these results, we derive the missing monthly values of the candidate indicators for each quarter in the forecast sample on the basis of the best auxiliary forecast models.

---

[12]The aggregate quarterly values of the indicators are calculated as the mean of the forecasted monthly series.

3.2.2 Real GDP forecasts with predicted monthly indicators

We generate quarterly forecasts of real GDP from the candidate indicators by readopting the recursive estimation procedure over the forecast sample from 2001Q1 to 2006Q3.[13] We implement the predictions of the monthly indicators that follow from the best auxiliary forecast models to construct the required quarterly values. For each indicator, Table 5 summarizes the results of the forecast performance tests, which are based on the one–quarter–ahead out–of–sample forecast errors.

Table 5: Forecast Properties at the Date of the WES Release

		RMSE	HLN–Test	TP–Test	p–value
Quantitative indicator					
Industrial production	IP	0.28	+0.41	14	0.11
Qualitative indicators					
CESifo Economic Climate	WES	0.22	−1.52	15	0.05
OECD indicator	OLI	0.22	−1.48	15	0.05
Economic sentiment	ESI	0.27	+0.27	15	0.05
EuroCOIN indicator	ECI	0.28	+0.76	12	0.34
Benchmark forecast					
AR(1) model	AR	0.26	−	13	0.20

Notes: For the HLN (1997) test the corresponding critical value is ± 1.31 for the 5% level with 22 degrees of freedom. A value of the HLN statistic below -1.31 implies an improvement, while a value above +1.31 implies a worsening of the forecast compared to the AR(1) benchmark prediction. TP denotes the number of correctly identified changes in the direction of real GDP growth; the p–value denotes statistical significance.

The forecast properties of the OLI clearly dominate those of the competing monthly indicators in terms of forecast accuracy. Only projections derived from the OLI outperform – as illustrated by the HLN (1997) test – the AR(1) benchmark forecast. In contrast the forecast performance of IP, the ESI and the ECI deteriorates considerably. In addition to the OLI, the ESI maintains the capacity of correctly predicting turning points.

For a comparison of the forecast properties of the WES with those of the competing monthly indicators, we employ the HLN (1997) test to evaluate the differences of the forecast errors for statistical significance. The results are shown

---

[13]Notice that the bridge models for the candidate indicators retain to those specifications that have been selected under full information.

in Table 6, which indicate that the WES surpasses industrial production, the ESI and the ECI unambiguously, while the OLI performs equally well.

Overall, the WES appears to constitute – in addition to the OLI – a comparable efficient forecast measure that is available at a relatively early moment in the quarter. Forecasts obtained from the WES dominate those derived from industrial production, the ESI and the ECI and improve upon the AR(1) benchmark forecast significantly. The poor performance of IP, the ESI and the ECI is – at least to some extent – attributed to the additional uncertainty arising from the necessity of extrapolating the missing monthly data.

Table 6: Forecast Comparison to the WES

	IP	OLI	ESI	ECI
HLN Statistic	+1.67	−0.05	+1.35	+1.49

Notes: HLN (1997) test of equal forecast performance of the WES and the competing monthly indicators. H_0 is rejected when the HLN statistic is above or below the critical value that amounts to ± 1.31 for the 5% significance level with 22 degrees of freedom.

The forecast performance of the OLI is comparatively strong since in contrast to the competing monthly indicators it does not deteriorate under incomplete information. Apparently for short–term forecasts of real GDP growth the OECD indicator seems to be an adequate measure, which can be relatively accurately extrapolated. Indeed, we find that an AR(2) process for the OLI captures the underlying time series properties in the sample period from 1991Q1 to 2006Q3 properly.

3.3 Real time evaluation of the forecast performance of the WES

Compared to competing monthly indicators and to univariate approaches the WES ensures a proper forecast performance concerning real GDP growth in the Euro area. However, this provides only limited comfort as one might be more interested in the forecast performance of a chosen model not only relative to an arbitrarily selected time series benchmark model but to forecasts of professional researchers and agencies. Yet, choosing the forecasts of a single agency is somehow again arbitrary and will reveal little in terms of the overall performance of the tested model, as they have different strengths and weaknesses over time and are thus difficult to rank. Due to diversification gains, combining a range of forecasts from professional agencies tends to outperform most individual predictions over

time and thus provides a fairly good benchmark for a chosen model.[14] In the following, we use the quarterly Consensus Forecasts for the Euro area published by Consensus Economics as point of reference. The Consensus Forecast is widely-used as a benchmark in the literature of out–of–sample forecasting and is well known as hard to beat. It is calculated as the arithmetic average of the individual predictions of the participating panelists. The quarterly Consensus Forecast for the Euro area is published only once a quarter, namely in the second week of the third month and is based on a survey in the previous two weeks.

Like many macroeconomic variables, real GDP growth is subject to data revisions as more accurate estimates become available. As the Consensus Forecast is built on an information set available at the time of publication, evaluating the predictions by means of today's revised real GDP time series and comparing their forecast abilities to those of the WES in this manner is somehow unequable and misleading. The use of real time data, i.e. vintage versions of data that were available on specific dates in history, for estimating and forecasting the chosen model specification and for calculating the forecast errors provides an adequate framework. The Euro Area Business Cycle Network (EABCN) provides vintage data of several macroeconomic variables for the Euro area in its EABCN Real Time Database (RTDB), based on series reported in the ECBs Monthly Bulletins.[15] To ensure comparability with the Consensus Forecasts as benchmark, we feed the specified bridge equation for the WES with vintage data of real GDP of the month of the WES release, which corresponds to the month when the first estimate of last quarter's GDP is published. We derive short-term forecasts of the current's quarter real GDP by adopting the described exercise of augmentation. The bridge model for the WES thereby retains the specification selected under full information.

Following Zarnowitz and Braun (1992) and Batchelor (2001) we use the values of real GDP available one year after the publication of the predictions as the relevant realizations for computing the forecast errors. Due to data limitations, our real time forecast horse race is restricted to 14 independent point forecasts.[16] As the quarterly Consensus Forecast for the Euro area is only updated in the last

---

[14] A large academic literature has studied the benefits of pooling forecasts from professional agencies. Batchelor and Dua (1995) showed that the Blue Chip Economic Indicators consensus forecasts for the US outperformed about 70–80 % of the panelists in the 1980s. Zarnowitz (1984) and McNees (1987) found similar results for a number of US macroeconomic variables as target.

[15] As the RTDB builds on the Euro area concept, the vintage data for real GDP comprises the EU12 and currently places quarterly time series on a monthly basis from January 2001 until December 2006 at the disposal.

[16] The quarterly Consensus Forecast for the Euro area is published only since the first quarter 2003. Following the procedure described above, we calculate forecast errors up to the predictions of the second quarter 2006.

month of the quarter, the comparison approach thus grants additional information of up to one month to the professional forecasters compared to the WES experts. This suggests that the forecast performance of the WES might be inferior. We evaluate the forecast properties of the WES and of the Consensus predictions by taking reference to the forecast accuracy tests. As the quarterly Consensus Forecasts for the Euro area are published as year-on-year growth rates, we convert the WES predictions to that unit in order to make both time series comparable. Table 7 summarizes the results of our real time forecast comparison.

Table 7: Real time evaluation of the forecast performance of the WES

	RMSE	HLN–Test
CESifo Economic Climate	0.32	0.77
Consensus Forecast	0.30	–

Notes: The HLN (1997) test is based on 14 independent point forecasts. The corresponding critical value for the 5% level is ± 1.35 with 13 degrees of freedom. A value of the HLN statistic below -1.35 implies a significant improvement, while a value greater that +1.35 implies a significant worsening of the forecast compared to the Consensus Forecast benchmark prediction.

Although the Consensus Forecast benefits from additional information of up to one month within the predicted quarter, it shows only a slightly lower out–of–sample forecast RMSE, but fails to outperform the WES in terms of real time out–of–sample forecast accuracy. This supports the results that the WES constitutes an accurate indicator in terms of deriving flash estimates of real GDP growth at a relatively early stage within the current quarter.

4   Conclusion

We have evaluated short–term forecasts of real GDP in the Euro area derived from the CESifo Economic Climate indicator (WES) in terms of forecast accuracy. The forecast properties of the WES have been compared to those of the ESI, OLI and the ECI. Considering the CESifo indicator is interesting because it differs from the monthly composite indicators in two specific aspects: (i) it is exclusively based on the assessment of economic experts about the current economic situation, and (ii) it is released within the quarter on a quarterly basis. A continuous monthly update of fresh monthly information within the survey quarter thus becomes impossible.

Our evaluation of the forecast performance of the WES has concentrated on both, the case of full information, which means that the competing monthly indi-

cators are completely known for the quarter, and on the case of incomplete information. The forecast sample has run from 2001Q1 to 2006Q3. Several forecast performance tests have been implemented, including tests on forecast accuracy and forecast directional correctness. Our findings have shown that the forecast power of the WES is comparatively proper.

Short–term forecasts of real GDP derived from the WES have the potential to provide an adequate understanding of the economic situation at an early moment in the quarter. This applies also to the OLI that has turned out to be the dominant composite indicator in terms of forecast accuracy. Comparing the forecast performance of the WES and Consensus Forecast by means of real time data supports the findings by showing that the rival predictions are equally precise.

Since the WES for the Euro area is also published for several member states it seems interesting to evaluate the forecast performance of the national indicators, which possibly provide a comprehensive insight on the current area–wide economic situation. Furthermore, short–term forecasts of real GDP derived from aggregate indicators are possibly outperformed by the aggregation of individual country forecasts derived from national indicators. Marcellino, Stock, and Watson (2003) find support for this conjunction by showing that forecasts constructed from the aggregation of individual country forecasts seem to be more accurate. As a consequence, comparing the forecast performance of the WES for the aggregate Euro area and the member states might be fruitful. In future research, these points will be addressed.

References

Altissima, F., A. Bassanetti, R. Cristadoro, M. Forni, M. Lippi, and L. Reichlin (2001): "EUROCOIN: a Real Time Coincident Indicator of the Euro Area Business Cycle," *CEPR Discussion Paper*, 3108.

Baffigi, A., R. Golinelli, and G. Parigi (2002): "Real Time GDP Forecasting in the Euro Area," Temi di Discussione 276, Banca d'Italia.

Banerjee, A., M. Marcellino, and I. Masten (2003): "Leading Indicators for Euro–Area Inflation and GDP Growth," Discussion Paper 3893, CEPR.

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

Batchelor, R., and P. Dua (1995): "Forecaster Diversity and the Benefits of Combining Forecasts," *Management Science*, 41(1), 68–75.

CAMBA-MENDEZ, G., R. G. KAPETANIOS, R. SMITH, AND M. R. WEALE (2001): "An Automatic Leading Indicator of Economic Activity: Forecasting GDP Growth for European Countries," *Econometric Journal*, 4, 856–890.

CLEMENTS, M. P., AND D. I. HARVEY (2006): "Forecast Encompassing Tests and Probability Forecasts," Warwick Economic Research Paper 774, University of Warwick.

DIEBOLD, F. X., AND R. S. MARIANO (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13(3), 253–263.

DIRON, M. (2006): "Short-term Forecasts of Euro Area Real GDP Growth – An Assessment of Real–Time Performance based on Vintage Data," Working Paper 622, ECB.

DOAN, T., R. LITTERMAN, AND C. SIMS (1984): "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Review*, 3, 1–100.

EUROPEAN COMMISSION (2007): "The Joint Harmonised EU Programme of Business and Consumer Surveys: User Guide," Economic studies and research, European Commission.

GAYER, C. (2005): "Forecast Evaluation of European Comission Survey Indicators," *Journal of Business Cycle Measurement and Analysis*, 2, 157–183.

GOLINELLI, R., AND G. PARIGI (2007): "The Use of Monthly Indicators to forecast Quarterly GDP in the Short–Run: An Application to the G7 Countries," *Journal of Forecasting*, 26, 77–94.

GRANGER, C. W. J. (1993): "On the limitations of comparing mean square forecast errors: Comment," *Journal of Forecasting*, 12(8), 651–652.

HARVEY, D. I., S. J. LEYBOURNE, AND P. NEWBOLD (1997): "Testing the Equality of Prediction Mean Squared Errors," *International Journal of Forecasting*, 13(2), 281–291.

MARCELLINO, M., J. STOCK, AND M. WATSON (2003): "Macroeconomic Forecasting in the Euro area: Country specific versus Area–Wide Information," *European Economic Review*, 47, 1–18.

MCNEES, S. K. (1987): "Consensus Forecasts: Tyranny of the Majority," *New England Economic Review*, pp. 15–21.

MOUROUGANE, A., AND M. ROMA (2003): "Can Confidence Indicators be useful to predict short–term real GDP Growth?," *Applied Economic Letters*, 10, 519–522.

OECD (2003): *Business Tendency Surveys: A Handbook.* OECD.

PARIGI, G., AND G. SCHLITZER (1995): "Quarterly Forecasts of the Italian Business Cycle by Means of Monthly Economic Indicators," *Journal of Forecasting*, 14, 117–141.

PESARAN, M. H., AND A. TIMMERMANN (1992): "A Simple Non–parametric Test of Predictive Performance," *Journal of Business and Economic Statistics*, 10(4), 461–465.

RÜNSTLER, G., AND F. SÉDILLOT (2003): "Short–Term Estimates of Euro Area Real GDP by Means of Monthly Data," Working Paper Series 276, European Central Bank.

SÉDILLOT, F., AND N. PAIN (2003): "Indicator Models of Real GDP Growth in Selected OECD Countries," Economics Department Working Papers 364, OECD.

STANGL, A. (2007): "World Economic Survey," in *Handbook Of Survey-Based Business Cycle Analysis*, ed. by G. Goldrian, pp. 57–67. Ifo Economic Policy Series, Edward Elgar.

ZARNOWITZ, V. (1984): "The Accuracy of Individual and Group Forecasts from Business Outlook Surveys," *Journal of Forecasting*, 3, 11–26.

ZARNOWITZ, V., AND P. BRAUN (1992): "Twenty–Two Years of the NBER–ASA Quarterly Outlook Surveys: Aspects and Comparisons of Forecast Performance," *NBER Working Paper*, 3965.

Ifo Working Papers

No. 45   Buettner, T., Reform der Gemeindefinanzen, April 2007.

No. 44   Abberger, K., S.O. Becker, B. Hofmann und K. Wohlrabe, Mikrodaten im ifo Institut – Bestand, Verwendung und Zugang, März 2007.

No. 43   Jäckle, R., Health and Wages. Panel data estimates considering selection and endogeneity, March 2007.

No. 42   Mayr, J. and D. Ulbricht, Log versus Level in VAR Forecasting: 16 Million Empirical Answers – Expect the Unexpected, February 2007.

No. 41   Oberndorfer, U., D. Ulbricht and J. Ketterer, Lost in Transmission? Stock Market Impacts of the 2006 European Gas Crisis, February 2007.

No. 40   Abberger, K., Forecasting Quarter-on-Quarter Changes of German GDP with Monthly Business Tendency Survey Results, January 2007.

No. 39   Batchelor, R., Forecaster Behaviour and Bias in Macroeconomic Forecasts, January 2007.

No. 38   Sülzle, K., Innovation and Adoption of Electronic Business Technologies, December 2006.

No. 37   Overesch, M. and G. Wamser, German Inbound Investment, Corporate Tax Planning, and Thin-Capitalization Rules – A Difference-in-Differences Approach, December 2006.

No. 36   Kempkes, G. and C. Pohl, The Efficiency of German Universities – Some Evidence from Non-Parametric and Parametric Methods, October 2006.

No. 35   Kuhlmann, A., German Productivity – A Reassessment via the New Ifo Productivity Database, October 2006.

No. 34   Kuhlmann, A., What is the X-Factor in the German Electricity Industry?, September 2006.

No. 33   Temple, J. and L. Wößmann, Dualism and cross-country growth regressions, August 2006.

No. 32 Baumann, F., V. Meier and M. Werding, Transferable Provisions in Individual Health Insurance Contracts, July 2006.

No. 31 Abberger, K., Qualitative Business Surveys in Manufacturing and Industrial Production – What can be Learned from Industry Branch Results?, May 2006.

No. 30 Ruschinski, M., Investigating the Cyclical Properties of World Trade, May 2006.

No. 29 Holzner, Chr., V. Meier and M. Werding, Time Limits in a Two-tier Unemployment Benefit Scheme under Involuntary Unemployment, April 2006.

No. 28 Eggert, W. and A. Haufler, Company Tax Coordination cum Tax Rate Competition in the European Union, April 2006.

No. 27 Lachenmaier, S. and H. Rottmann, Employment Effects of Innovation at the Firm Level, April 2006.

No. 26 Radulescu, D.M. and M. Stimmelmayr, Does Incorporation Matter? Quantifying the Welfare Loss of Non-Uniform Taxation across Sectors, March 2006.

No. 25 Lessmann, Chr., Fiscal Decentralization and Regional Disparity: A Panel Data Approach for OECD Countries, March 2006.

No. 24 Fuchs, Th., Industry Structure and Productivity Growth: Panel Data Evidence for Germany from 1971–2000, December 2005.

No. 23 Holzner, Chr. and A. Launov, Search Equilibrium, Production Parameters and Social Returns to Education: Theory and Estimation, December 2005.

No. 22 Sülzle, K., Stable and Efficient Electronic Business Networks: Key Players and the Dilemma of Peripheral Firms, December 2005.

No. 21 Wohlrabe, K. and M. Fuchs, The European Union's Trade Potential after the Enlargement in 2004, November 2005.

No. 20 Radulescu, D.M. and M. Stimmelmayr, Implementing a Dual Income Tax in Germany: Effects on Investment and Welfare, November 2005.