Ifo Institute for Economic Research at the University of Munich

ifo Working Papers

Testing Predicitive Ability of Business Cycle Indicators for the Euro Area

Christina Ziegler

Ifo Working Paper No. 69

March 2009

Testing Predicitive Ability of Business Cycle Indicators for the Euro Area

Abstract

We analyze the predictive power of seven leading indicators for economic activity in the Euro Area developed by different banks, institutions and research centers. Our comparison is conducted in a bivariate vector autoregressive framework. Indicators are compared by means of an in-sample and an out-of-sample forecasting experiment. Predictive accuracy is compared by recently proposed tests for superior predictive ability. Our results suggest that nearly all indicators have good in-sample properties and that a majority of them is able to outperform a naive univariate autoregressive model out-of-sample. Additionally, we find that indicators perform better in boom periods than in recessions. The *OECD* and *FAZ* indicators are both composite indicators and deliver the most accurate forecasts.

JEL Code: C32, C53, E32.
Keywords: Leading indicators, Euro area, forecasting.

Christina Ziegler
Ifo Institute for Economic Research
at the University of Munich
Poschingerstr. 5
81679 Munich, Germany
Phone: +49(0)341/9733565
ziegler@ifo.de

I would like to thank an anonymous referee, Robinson Kruse, Jörg Breitung, Jonas Dovern, Arne Gieseck and the participants of the workshop „Makroökonomik und Konjunktur“ at the Ifo Institute Dresden as well as of the Ph.D. seminar in Leipzig for inspiring discussions and helpful comments.

1. Introduction

The analysis of business cycles and its characteristics has a long tradition in economic research. In addition, economic growth is beside unemployment and inflation one of the most important target variables in the decision process of policy makers. Due to increasing political and economical uncertainty about future developments, reliable forecasts become more and more important.

In our work, we concentrate on business cycle indicators for the Euro Area that have been developed by research institutes, banks and the European Commission in order to improve forecasts and reduce uncertainty. These indicators are important tools for enterprizes, central bankers and politicians to predict the future development of the economy. We analyze the predictive ability of seven indicators which are quite different regarding their conception. The empirical analysis shows that they also have significantly different forecasting performances. In contrast to related articles in this area we consider seven special economic indicators that are used in practice to conduct economic forecasting. These indicators are constructed with a specific focus on the European business cycle movements. However, the forecasting abilities of these indicators have not been analyzed so far from a comparative perspective to the best of our knowledge. Our comparison is conducted in two ways: the in-sample and the out-of-sample analysis. The former uses all available information to estimate cross-correlations and to test against Granger causality. The latter tries to mimic a realistic situation where the future is unknown. We make use of the bivariate vector autoregressive framework to generate one-step ahead forecasts of year-over-year growth of industrial production which serves as the reference series. Our choice is motivated by the fact that this series is available at a monthly frequency implying a larger number of observations compared to the quarterly GDP series.

A great deal of literature concentrates on the forecasting abilities of other economic indicators such as the interest rate spread or indicators such as M2 growth (see Dovern and Ziegler (2008)). At forecasting real GDP our paper is related to rather extensive literature that assesses the forecasting properties of various leading indicators for Germany such as by Breitung and Jagodzinski (2001), Schumacher and Dreger (2004), Fritsche and Stephan (2002), Hüfner and Schröder (2002), Schumacher (2007) or Kholodilin and Siliverstovs (2006) and the Euro Area such as Forni et al. (2003).

This paper is organized as follows. Section 2 presents the data base and especially the components of different indicators. Section 3 covers the methods and results of the in-sample analysis, while section 4 is dedicated to an out-of-sample experiment and recent tests for superior predictive ability. Main conclusions are drawn in section 5.

2. Database

2.1. Reference Series

Our reference series is given by the year-over-year (yoy) growth rate of industrial production index for the Euro Area (source: Eurostat), which is available at a monthly frequency. Alternatively, we may have used the quarterly recorded GDP but this would imply a much smaller sample size. It is a well known fact that estimators perform better and inference gets more reliable as the sample size increases. Therefore, we employ the industrial production series as reference. Although this production-index counts only one third of the total GDP, most impulses leading the business cycle are caused by the industrial sector (see Breitung and Jagodzinski (2001)). In addition, we implement a balanced sample size across indicators, meaning that the data spans from *1991M02* to *2007M07*.

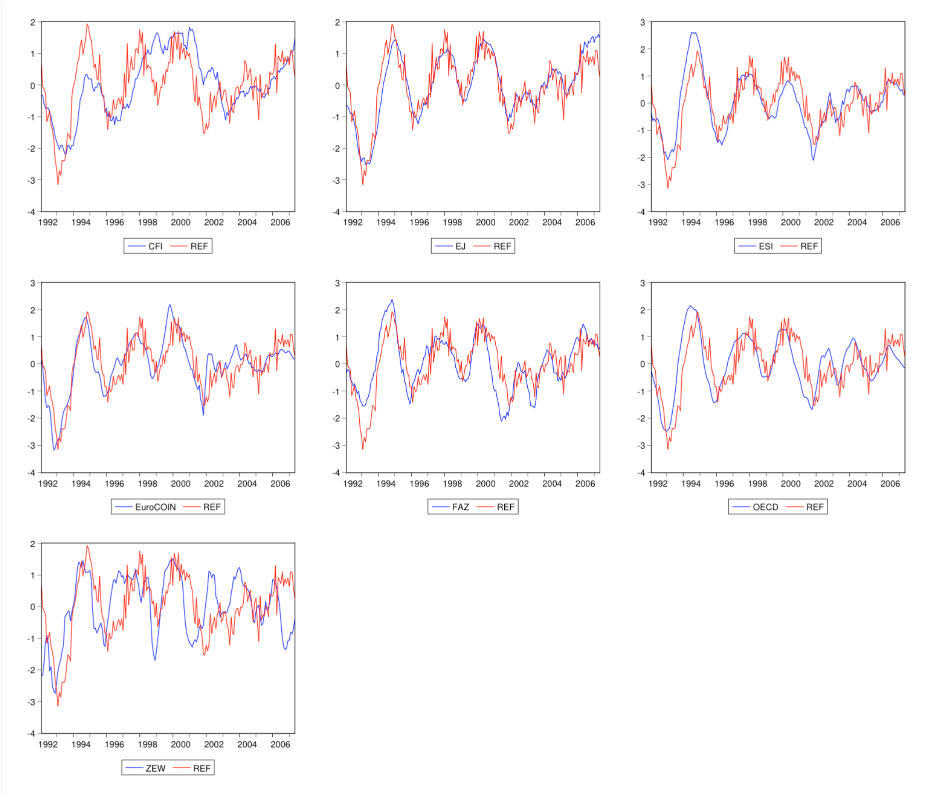

Figure 1: Indicator series

2.1. Business Cycle Indicators

We consider in our study seven different business cycle indicators for the Euro Area. These indicators are constructed from different newspapers, banks or economic research centers to predict or show future trends of the European business cycle. The indicators are selected on the basis of data availability. These different types of indicators could be clustered more or less into three different groups. Indicators belonging to the first class are constructed with the help of different surveys. These indicators are constructed and influenced by different surveys evaluated by an institution. The European Sentiment Indicator (*ESI*) and the Business and Climate Indicator (*EJ*) are constructed by the European Commission and conclude several survey results evaluated by the Commission. The *ESI* indicator contains five different confidence indicators, where different sectors are asked for their expectation of the European Business cycle. Analogous to other studies (see Kholodilin and Siliverstovs (2006)), we decide to include the consumer confidence indicator, which is also a part of the *ESI* indicator, also separately. The *EJ* indicator resembles the European business climate. This indicator contains elements of the industry confidence indicator, also included in the *ESI* plus other survey results. Next to the indicators evaluated by European Commission we analyze a business cycle indicator, proposed by the *ZEW*, which is also constructed throughout surveys, where about 350 different business and financial experts are asked. The survey deals with the markets of Germany, the United States, Japan, Great Britain, France and Italy[1].

The second category includes composite indicators. These indicators contain different time series, which are assumed to have explanatory power and leading abilities for the European business cycle, e.g. job vacancies, interest rates etc.. This group covers the *FAZ* indicator, developed by the DZ bank and the business cycle indicator proposed by the OECD (*OECD*).

The last group of indicators contains only the EuroCoin (*EC*) indicator, which is constructed by a forecast based on dynamic factor models (see Forni et al. (2003) and Altissimo et al. (2006)) and is monthly published from the *CEPR*. In this indicator factors

---

[1] For the ZEW indicator, data for the Euro Area are first available from 1999 on, we decide to use *ZEW* survey results for Germany before 1999 as a proxy. The correlation between these two series is approximately 0.85.

from more than one hundred different time series from eleven different categories are extracted with a principal component analysis and are then used for the construction of the *EuroCOIN*. Table 1 contains a list of the indicators and their compositions. Figure 1 displays the standardized indicators and reference series.

Table 1: Overview of different indicators

Indicator	**Components**	**Source**
European Sentiment Indicator (ESI)	Industry Confidence Indicator, Services Confindence Indicator Consumer Confidence Indicator (CFI) Construction Confidence Indicator Retail Trade Confidence Indicator	European Commission
Business Climate Indicator (EJ)	Industry survey about: production trends in recent months, order books export order books, stocks and production expectations	European Commission
FAZ-Euro-Indicator (FAZ)	New job vacancies, order entries, Reuter purchasing manager´s index (PMI), building and planning permissions, production, interest rate spread, cosumer confidence, Morgan-Stanley- Capital-International Index, real money (M3)	DZ-Bank
OECD Composite Indicator (OECD)	Composite by individual OECD indicators for EU-12: variables for surveys by national institutes, new job vacancies, orders inflow/demand, spread of interest rates, production, finished goods stocks, passenger car registration, other national indicators	Organisation for Economic Co-operation and Development (OECD)
ZEW Indicator of Economic Sentiment (ZEW)	Medium-term expectations for development of the macroeconomic trend, inflation rate, short-term and long-term interest rates, stockmarket, exchange rates, profit situation of different German industries (only financial experts)	Centre for European Economic Research (ZEW)
EuroCoin (EC)	Data from 11 categories: industrial production, producer prices, monetary aggregates, interest rates, financial variables, exchange rates, surveys by the European Commission, surveys by national institutes, external trade, labour market	Centre for Economic Policy Research (CEPR)

3. In-Sample-Analysis

We compute cross-correlations for various lags and leads and test against Granger causality within a bivariate vector autoregressive framework. Before doing so, we apply the stationarity test by Kwiatkowski et al. (1992) [KPSS] to confirm the stationarity of considered time series. It is a well known fact that non-stationary time series has to be handled with a special care.

3.1. Stationarity Test

The time series are displayed in Figure 1. Some of them show upward trending behavior and appear to be stationary which is not surprising since we consider growth rates. In order to test the hypothesis of stationarity formally, we apply the KPSS stationarity test in our preliminary data analysis. Results of the KPSS test that is applied to de-meaned (denoted as *c* in Table 2) or de-trended data (denoted as *c, t)* can be found in Table 2. The overall impression is that nearly all time series are stationary because the null hypothesis has to be rejected in only one case (*CFI*). Due to the fact that we use asymptotic critical values that might not be exact in small samples and the circumstance that the *CFI* series does not look quite different from all others, we decide to treat it as stationary. In addition, it does not make much sense to have a non-stationary predictor for a stationary reference series[2].

Table 2: KPSS Test Results

REF	CFI	EC	EJ	ESI	FAZ	OECD	ZEW
0.104	0.232	0.137	0.111	0.061	0.087	0.074	0.215
c,t	c,t	c,t	c,t	c	c	c	c
$I(0)$	$I(1)$	$I(0)$	$I(0)$	$I(0)$	$I(0)$	$I(0)$	$I(0)$

Notes: Inference about the order of integration is drawn at the five percent level of significance. c and c,t corresponds to de-meaned and de-trended data, respectively. The automatic bandwidth of the Bartlett kernel estimator is determined by the Newey-West (1994) method. Inference is drawn at the ten percent level of significance.

[2] An application of Johansen's cointegration test does not contradict the assumption of stationarity. This holds for the trace statistic as well as for the maximum eigenvalue statistic.

3.2. Cross Correlation Analysis

The cross-correlation analysis can be considered to investigate the in-sample predictive power of indicators. The cross-correlation coefficient for lag $k=-12,\ldots,0,\ldots,12$ is given by the correlation coefficient between the reference series y_t and the indicator x_{t-k}. If the indicator is leading the reference series then the cross-correlation should be as high as possible for positive values of k. The results are shown in Table 3. The *CFI* and *EJ* indicator are not leading in the sense that their maximal cross-correlation coefficient is associated with a negative value of k. On the contrary, other indicators have satisfying properties. The *ZEW* indicator reaches its highest correlation (0.696) at $k=5$ suggesting that it led the reference series for five months.

Table 3: Cross Correlation Analysis

	CFI	EC	EJ	ESI	FAZ	OECD	ZEW
max	0.697	0.883	0.906	0.828	0.801	0.868	0.696
k	-4	2	-1	2	3	4	5

Notes: max refers to the maximal estimated cross-correlation coefficient at lag k.

3.3. Granger Causality Test

The Granger causality test (Granger (1969)) is used to check whether an indicator x_t can improve the prediction of the reference series y_t. To put it differently, we ask if lagged values of x_t exhibit some prediction ability for y_t . To perform this test, a VAR model of order p is used:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \ldots\ldots + \alpha_p y_{t-p} + \beta_1 x_{t-1} + \ldots\ldots + \beta_p x_{t-p} + \varepsilon_{1,t} \tag{1}$$

$$x_t = \gamma_0 + \gamma_1 y_{t-1} + \ldots\ldots + \gamma_p y_{t-p} + \delta_1 x_{t-1} + \ldots\ldots + \delta_p x_{t-p} + \varepsilon_{2,t}, \tag{2}$$

where $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are assumed to be white noise processes. The null hypothesis can be expressed as parameter restrictions in the VAR model, i.e. $H_0: \ \beta_1 = \ldots = \beta_p = 0$. Table 4 contains the results of the Granger causality test. Since all p-values are zero we conclude that all indicators have a significant in-sample predictive power.

Table 4: Granger Causality Test Results

CFI	EC	EJ	ESI	FAZ	OECD	ZEW
0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: Given values are p-values for the null hypothesis of no causality. The lag length in the bivariate VAR model is chosen by Schwarz criterion.

4. Out-of-Sample Analysis

The VAR model described in the previous section is a standard modeling framework in macroeconomic forecasting. It is applied in this work in order to conduct one-step out-of-sample forecasts using the seven different leading indicators. It is well known that the linear VAR model is simple to estimate and therefore widely applied. On the contrary, it is not very flexible but previous evidence has shown that it is able to produce competitive forecasts.

In our forecasting experiment the consistent Schwarz criterion is used in each forecasting step, to choose the optimal lag length in the VAR model. We select an autoregressive model (AR) as our benchmark. Analogous to our VAR approach, the optimal lag length of the autoregressive terms is chosen with the Schwarz criterion. For both cases the maximum lag length is 15, which allows for rich dynamics. Furthermore we use a rolling window forecasting scheme to obtain different indicator forecasts. A forecast based on a rolling scheme relies on a fixed-length window which is shifted every period and the model and its lag length is re-estimated. The estimation sample period covers approximately ten years of our whole sample, from *1991M2* till *2001M12*. The number of observations in the first estimation time period is equal to the size of our window length in the rolling window forecasting scheme (119). Our forecasting period starts with the real introduction of the Euro at *2002M1* and ends five years later at *2007M7*. In Figure 2 the individual indicator and autoregressive forecasts are compared with the reference series.

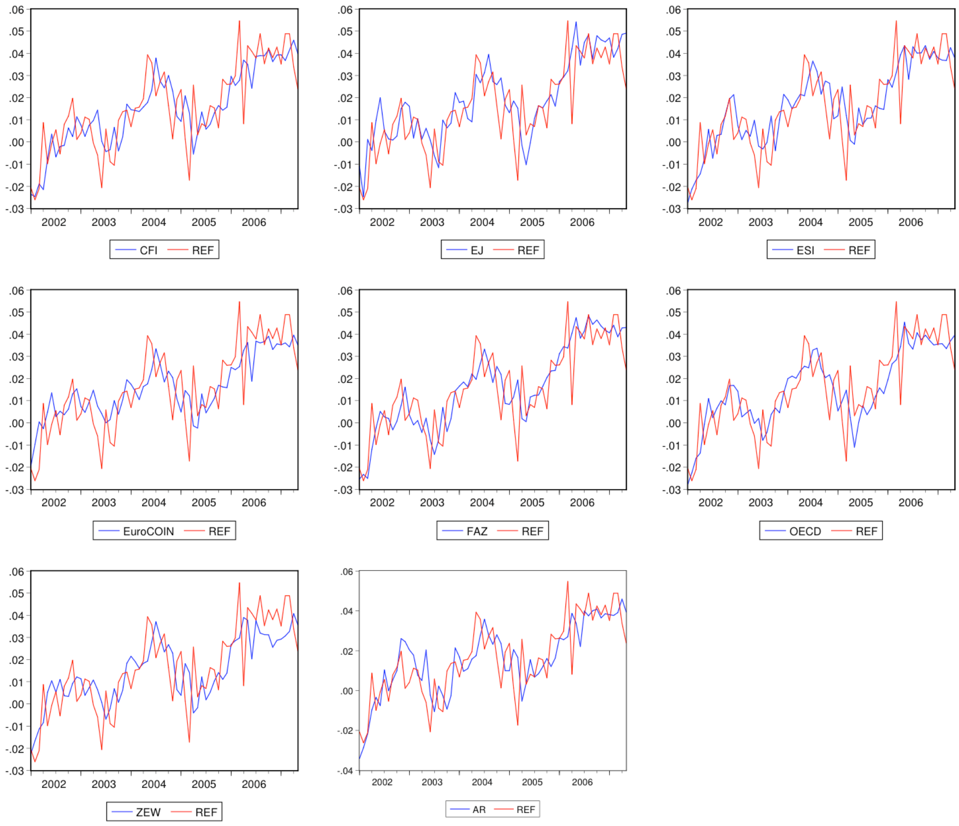

Figure 2: Individual indicator and autoregressive forecasts

4.1. Simple Evaluation Criteria

We compute for all our forecasts relative mean square errors (MSEs) and relative mean absolute errors (MAEs), see Table 5a and 5b (unweighted):

$$MSE = \frac{1}{P}\sum_{t=1}^{P}\left(y_t - y_t^f\right)^2 \tag{3}$$

$$MAE = \frac{1}{P}\sum_{t=1}^{P}\left|y_t - y_t^f\right| \tag{4}$$

Table 5a: Forecast Evaluation								
Rel. MSE	CFI	EC	EJ	ESI	FAZ	OECD	ZEW	AR
Unweighted	0.927	0.939	0.839	0.892	0.779	**0.760**	0.991	1.000
Rank	5	6	3	4	2	1	7	8
Left Tail	0.882	1.060	0.883	0.886	0.816	**0.768**	0.911	1.000
Rank	3	8	4	5	2	1	6	7
Right Tail	0.981	0.793	0.787	0.901	**0.734**	0.750	1.089	1.000
Rank	6	4	3	5	1	2	8	7
Rel. MAE	CFI	EC	EJ	ESI	FAZ	OECD	ZEW	AR
Unweighted	0.947	0.967	0.907	0.966	**0.861**	0.885	1.049	1.000
Rank	4	6	3	5	1	2	8	7
Left Tail	0.917	1.045	0.938	0.984	**0.881**	0.899	1.007	1.000
Rank	3	8	4	5	1	2	7	6
Right Tail	0.980	0.882	0.874	0.947	**0.839**	0.869	1.095	1.000
Rank	6	4	3	5	1	2	8	7

Table 5b: Tails relative to Full distribution								
Rel. MSE	CFI	EC	EJ	ESI	FAZ	OECD	ZEW	AR
Left Tail	0.522	0.619	0.577	0.545	0.575	0.554	0.504	0.549
Rank	2	8	7	3	6	5	1	4
Right Tail	0.478	**0.381**	0.423	0.455	0.425	0.446	0.496	0.451
Rank	7	1	4	3	3	5	8	6
Rel. MAE	CFI	EC	EJ	ESI	FAZ	OECD	ZEW	AR
Left Tail	0.506	0.564	0.540	0.532	0.534	0.531	0.501	0.522
Rank	2	8	7	5	6	4	1	3
Right Tail	0.494	**0.436**	0.460	0.468	0.466	0.469	0.499	0.478
Rank	7	1	2	3	4	5	8	6

Notes (5a): Rel. MSE and Rel. MAE denote the MSE and MAE of an indicator-based forecast relative to the autoregressive benchmark forecast, respectively. Left Tail and Right Tail correspond to weighted forecast errors, see (7) and (8), respectively. Values smaller than one indicate better performance than the benchmark. Bold values indicate the best performance.

Notes (5b): Rel. MSE and Rel. MAE denote the MSE and MAE of an indicator-based forecast with weighted errors relative to its MSE and MAE, respectively. Left Tail and Right Tail correspond to weighted forecast errors, see (7) and (8), respectively. Values smaller than one-half indicate better performance than the benchmark. Bold values indicate the best performance.

where y_t is the realized value, y_t^f the predicted value and P is the number of out-of-sample observations (*P=65*) in the individual forecast.

Let us denote the relative MSEs and MAEs as:

$$\theta = \left(\frac{MSE^{Indicator}}{MSE^{AR}} \right) \tag{5}$$

$$\delta = \left(\frac{MAE^{Indicator}}{MAE^{AR}} \right) \tag{6}$$

where *Indicator* denotes the indicator based forecast and *AR* the benchmark autoregressive forecast. If $\theta < 1$ or $\delta < 1$ the indicator forecast has a superior forecasting performance (smaller MSE (MAE)) than the benchmark model and vice versa. Note that all forecast errors have the same weight in MSE and MAE, henceforth we call them in the following unweighted. Since accurate forecasts of extreme observations that are located in the tails of the reference series' distribution might be more important for policy makers than observations that are located around the mean, which is in the neighborhood of zero, we consider weight functions proposed by van Dijk et al. (2003) . In particular they are designed to put relatively more weight on forecast errors in times of more extreme observations, i.e. booms or recessions. Formally, the weight functions for the left tail (recession periods) and right tail (boom periods) are given by

$$L(y_t - y_t^f) = \left(1 - \hat{F}(y_t)\right)\left(y_t - y_t^f\right) \tag{7}$$

$$R(y_t - y_t^f) = \hat{F}(y_t)\left(y_t - y_t^f\right), \tag{8}$$

where $\hat{F}(y_t)$ denotes the empirical cumulative density function y_t and $(y_t - y_t^f)$ is the forecast error in period *t*. Figure 3 depicts the empirical cumulative density function of y_t.

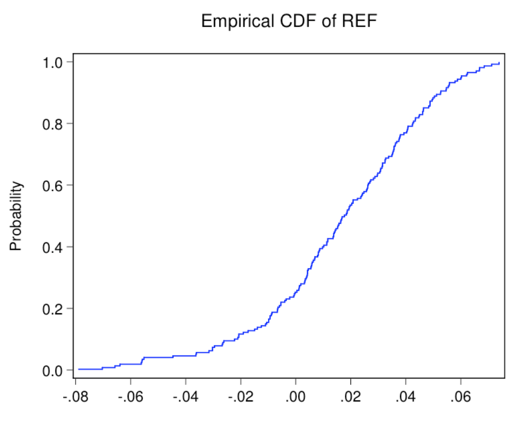

Figure 3: Empirical Cumulative Density Function $\hat{F}(y_t)$

In Table 5a we report computed MSE and MAE values of indicator forecasts relative to the corresponding values of the benchmark forecast. In contrast, Table 5b shows the performance of individual forecasts in both tails (weighted) relative to their performance over the full distribution (unweighted). Note that values smaller than 0.5 indicate better performance in the particular tail and vice versa. The threshold of 0.5 instead of 1 for unweighted loss functions is due to the fact that the means of $\hat{F}(y_t)$ and $\left(1-\hat{F}(y_t)\right)$ are equal to 0.5.

The results displayed in Table 5a show that, with the exception of the *ZEW*, that all indicators beat the *AR* model when comparing the relative MAEs and MSEs. The *FAZ* indicator delivers the smallest relative MAE as well as in the unweighed forecast and promises to be the best suited in forecasting booms and recessions when looking at the the relative MAE. The *OECD* indicator leads also to good results. When comparing the relative MSEs of all forecasts, in our observed time period this indicator was able to predict recessions best. The results reported in Table 5b show that all forecasts are more accurate in the right tail. The *EC* indicator in particular, improves much in the right tail compared to its performance over the full distribution.

All in all, we conclude that the indicators do much better in forecasting booms than recessions. At forecasting recessions some indicator forecasts are beaten by the autoregressive model. The *OECD* indicator performs best when forecasting recession, while the *FAZ* indicator has the smallest relative MSE and MAE at forecasting booms. Compared to the full distribution, the EC performs much better at forecasting booms than recessions.

4.2. Testing Rationality

In this section we apply the Rationality Test proposed by Mincer and Zarnowitz (1969) and Zarnowitz (1985). This test is used to analyze if the forecast error is uncorrelated with the forecast itself and the forecast is unbiased. If there is any correlation, important information has not been incorporated. The rationality test is based on the following test regression:

$$y_{t+1} = \alpha + \beta y^f_{t+1} + u_{t+1}, \tag{9}$$

where we assume that u_{t+1} is a white noise process. The null hypothesis can be formulated as $H_0:\ \alpha = 0, \beta = 1$. If the null hypothesis is rejected in favor of the alternative, we conclude that the forecast is not rational. However, the coefficient of determination R^2 provides a direct assessment of the variability in the reference series that is explained by the forecasts. Therefore it is often interpreted as a simple measure of the degree of predictability in the reference series, and hence of the potential economic significance of the forecasts.

When applying the Mincer-Zarnovitz regression to our forecasting errors derived from the individual indicator forecasts, we summarize that nearly all forecasts are rational and that no important information has been lost. On the contrary, the null hypothesis has to be rejected for the *FAZ* and the *EC* forecasts at a significant level of ten percent. Although the *FAZ* forecast is correlated somewhat with the contemporaneous forecast error, it delivers the highest value for R^2. Only the *ZEW* indicator exhibits less predictive power (in terms of the coefficient of determination) than the benchmark.

Table 6: Mincer-Zarnowitz Regression Results

	CFI	EC	EJ	ESI	FAZ	OECD	ZEW	AR
p-value	0.184	0.020	0.297	0.754	0.042	0.689	0.834	0.115
R^2	0.570	0.593	0.604	0.567	**0.655**	0.632	0.517	0.542

Notes: Reported p-values are computed for the null hypothesis $H_0 : \alpha = 0, \beta = 1$ in (9). R^2 denotes the coefficient of determination in the test regression (9).

4.3. Testing Superior Predictive Ability

In this work we are faced with several potential leading indicators that we use for forecasting purposes of one time series. Orthodox econometric techniques for forecast evaluating focus on testing pair-wise equal predictive ability of available forecasts. Beside the fact that multiple testing problems may arise, we aim at testing superior predictive ability. In other words, we would like to test the hypothesis that there is no indicator-based forecast outperforming the benchmark model by taking into account all forecasts simultaneously. We therefore make use of a recently proposed test by Hansen (2005) which is a modification of White´s (2000) seminal work in two respects which are discussed after introducing the test formally.

Let $X_{t,k} \equiv l_{t,0} - l_{t,k}$ be the loss difference between a benchmark (labeled as *0*) and another forecast (labeled as *k=1,...,l*). Note, that we employ $l_{t,k} = (y_t - y_{t,k}^f)^2$ (MSE) or $l_{t,k} = \left| y_t - y_{t,k}^f \right|$ (MAE) and that the number of elements in the forecast model set M_c equals *l+1*. If $E(X_{t,k}) > 0$ holds, we conclude that the benchmark is worse performing than the forecast *k*. On the contrary, if $E(X_{t,k}) \le 0 \; \forall k$, then the benchmark is (weakly) superior to all alternative forecasting models that are element of M_c In this case the benchmark exhibits superior predictive ability. We choose the autoregressive forecast as the benchmark, see section 4. A formal expression of the multiple null hypothesis stated above is given by

$$H_0 : \max_k E(X_{t,k}) \leq 0 \tag{10}$$

In other words, no forecast $k \in M_c$ is better than the benchmark model in terms of the particular loss function. The maximum operator takes into account that the maximal expected loss difference is the most relevant. If the null hypothesis is rejected, we conclude that there is at least one forecasting model that is superior to the benchmark. Of course the expectation of $X_{t,k}$ is unexpected but can be consistently estimated with the sample mean $\overline{X}_k = \frac{1}{P}\sum_{t=R+1}^{T} X_{t,k}$, where *k=1,...,l*. White (2000) proposed the following test statistic for the null hypothesis in equation (10):

$$t = \max_k P^{1/2}\overline{X}_k \tag{11}$$

Note that the limiting distribution of *t* is not unique under the null hypothesis. Therefore the stationary bootstrap method of Politis and Romano (1994) is utilized to obtain the distribution of *t* which we label *t**. This way of bootstrapping the distribution builds on randomly drawing subsamples from independent lengths which are drawn from a geometric distribution with mean *q*. These subsamples are randomly put together to obtain a bootstrapped series. This procedure is repeated *B* times.

There are two main problems with this approach that are commented in Bao et al. (2006). First, the choice of the forecasting scheme is not irrelevant. Nevertheless, a recursive scheme is also quite attractive, but the bootstrap method of Politis and Romano (1994) requires a special assumption that cannot be reconciled with such a forecasting scheme, see Hansen (2005). To be more precise, old observations have a higher probability to be drawn than newer ones in a recursive setting which is avoided in a rolling window setting. Second, the Reality Check test of White (2000) is conservative and depends heavily on the structure of M_c. If this set contains poor forecasts then White's test is conservative, since it assumes that all competing forecasting models are precisely as good as the benchmark.

Table 7: SPA Test Results

MSE	Unweighted	Left Tail	Right Tail
Upper	0.032	0.084	0.025
Consistent	0.032	0.089	0.025
Lower	0.032	0.089	0.025
MAE	**Unweighted**	**Left Tail**	**Right Tail**
Upper	0.065	0.182	0.043
Consistent	0.065	0.192	0.043
Lower	0.065	0.192	0.043

Notes: Reported values are p-values for the multiple null hypothesis that the benchmark is not outperformed by any other competing forecast. Upper, Consistent and Lower refer to different p-values explained in the text. Left Tail and Right Tail correspond to weighted forecast errors, see (7) and (8), respectively.

A solution to the last problem can be found in Hansen (2005), where a standardized test statistic is proposed. This test statistic is given by

$$\tilde{t} = \max_k \frac{P^{1/2}\overline{X}_k}{\hat{\omega}_{kk}} \tag{12}$$

where $\hat{\omega}_{kk}$ is an estimator of the asymptotic standard deviation of $P^{1/2}\overline{X}_k$. In order to avoid White's assumption that makes the test conservative, a different way of bootstrapping the distribution of $\tilde{t}$ was proposed, for details see Hansen (2005).

In addition, two inconsistent probability values can be provided in order to obtain a lower and an upper bound for the consistent probability value. The upper bound corresponds to the probability value of White's Reality Check test that is conservative. The lower bound corresponds to the probability value of a liberal test whose null hypothesis assumes that models with worse performance than the benchmark are poor models in the limit, see Hansen (2005).

Table 7 contains the results for our application. When looking at the results for the MSE we find that the null hypothesis is rejected in each case for a given significance level of ten percent. We observe that the p-values are close together which means that there are no relatively poor forecasts in our model set. Turning to the MAE, we have to change our conclusion in one case with respect to the relative performance of the benchmark model in periods of recessions: the indicator forecasts are not able to outperform the benchmark significantly.

5. Conclusions

In this paper we investigate the predictive performance of seven leading indicators for the Euro Area. Using the VAR framework we find that indicators have significantly high in-sample predictive ability. After having generated out-of-sample forecasts for the period after the introduction of Euro cash money in January 2002 we apply established as well as recent techniques to evaluate their predictive power. We focus on the performance during periods of booms and recessions by applying weight functions proposed by van Dijk et al. (2003). Our findings suggest that the performance of indicators varies with the business cycle. This result is underlined by an application of Hansen's test against superior predictive ability, see Hansen (2005).

Our results of the out-of-sample experiment imply that most indicators are able to beat a univariate autoregressive benchmark model in terms of popular loss functions like the mean squared error. Additionally, we find that a lot of indicators perform much better in boom-periods than in times of recessions. The *OECD* and *FAZ* (both composite indicators) deliver the most accurate VAR forecasts. The forecast with the *EC* indicator, which is based on a dynamic factor model, performs best in the times of booms. In times of recessions no indicator is able to beat the forecast by the simple autoregressive model.

References:

Altissimo, F., R. Cristadoro, M. Forni, M. Lippi and G. Veronese (2006), “New EuroCOIN: Tracking Economic Growth in Real Time,” Technical Report 5633, CEPR Discussion Paper.

Bao, Y. , T.-H. Lee and B.Saltoglu (2006), “Evaluating Predictive Performance of Value-at-Risk Models in Emerging Markets: A Reality Check,” *Journal of Forecasting*, 25, 101–128.

Breitung, J. and D. Jagodzinski (2001), “Prognoseeigenschaften alternativer Indikatoren für die Konjunkturentwicklung in Deutschland,” *Konjunkturpolitik*, 47, 292–314.

Dovern, J. and C. Ziegler (2008), "Predicting Growth Rates and Recessions. Assessing U.S. Leading Indicators under Real-Time Conditions", Applied Economics Quarterly 54, 293-318.

Dick, van D. and P. H. Franses (2003), “Selecting a Nonlinear Time Series Model using Weighted Tests of Equal Forecast Accuracy,” *Oxford Bulletin of Economics and Statistics*, 65, 727–744.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2003), “Do financial variables help forecasting inflation and real activity in the euro area?” *Journal of Monetary Economics*, 50, 1243–1255.

Fritsche, U. and S. Stephan (2002), “Leading Indicators of German Business Cycles: An Assessment of Properties”, *Journal of Economics and Statistics*, 223, 289-315.

Granger, C. W.J. (1969), “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods,” *Econometrica*, 37, 424–438.

Hansen, P. R. (2005), “A Test for Superior Predictive Ability,” *Journal of Business and Economic Statistics*, 23, 365–380.

Hüfner, F. P. and M. Schröder (2002), “Forecasting German Industrial Production: An Econometric Comparison of ifo- and ZEW-Business Expectations,” *Journal of Economics and Statistics*, 222, 316–336.

Kholodilin, K. A. and B. Siliverstovs (2006), “On the Forecasting Properties of the Alternative Leading Indicators for the German GDP: Recent Evidence,” *Jahrbücher für Nationalökonomie und Statistik*, 226(3), 234–259.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt and Y. Shin (1992), “Testing the null hypothesis of stationarity against the alternative of a unit root. How sure are we that economic time series have a unit root?” *Journal of Econometrics*, 54, 159–178.

Mincer, J., and V. Zarnowitz (1969), „The Evaluation of Economic Forecasts“ in J. Mincer (ed.), Economic Forecasts and Expectations, National Bureau of Economic Research, New York.

Politis, D. N., and J. P. Romano (1994), “The Stationary Bootstrap,” *Journal of the American Statistical Association*, 89, 1303–1313.

Schumacher, C. (2007), “Forecasting German GDP Using Alternative Factor Models Based on Large Datasets,” *Journal of Forecasting*, 26, 271–302.

Schumacher, C. and C. Dreger (2004), “Estimating large-scale factor models for economic activity in Germany: Do they outperform simpler models?” *Jahrbücher für Nationalökonomie und Statistik*, 224, 731–650.

White, H. (2000), “A reality check for data snooping,” *Econometrica*, 68, 1097– 1126.

Zarnovitz, V. (1985), “Rational expectations and macroeconomic forecasts,” *Journal of Business and Economic Statistics*, 3, 293–311.

Ifo Working Papers

No. 68 Schütz, G., Does the Quality of Pre-primary Education Pay Off in Secondary School? An International Comparison Using PISA 2003, March 2009.

No. 67 Seiler, C., Prediction Qualities of the Ifo Indicators on a Temporal Disaggregated German GDP, February 2009.

No. 66 Buettner, T. and A. Ebertz, Spatial Implications of Minimum Wages, February 2009.

No. 65 Henzel, S. and J. Mayr, The Virtues of VAR Forecast Pooling – A DSGE Model Based Monte Carlo Study, January 2009.

No. 64 Czernich, N., Downstream Market structure and the Incentive for Innovation in Telecommunication Infrastructure, December 2008.

No. 63 Ebertz, A., The Capitalization of Public Services and Amenities into Land Prices – Empirical Evidence from German Communities, December 2008.

No. 62 Wamser, G., The Impact of Thin-Capitalization Rules on External Debt Usage – A Propensity Score Matching Approach, October 2008.

No. 61 Carstensen, K., J. Hagen, O. Hossfeld and A.S. Neaves, Money Demand Stability and Inflation Prediction in the Four Largest EMU Countries, August 2008.

No. 60 Lahiri, K. and X. Sheng, Measuring Forecast Uncertainty by Disagreement: The Missing Link, August 2008.

No. 59 Overesch, M. and G. Wamser, Who Cares about Corporate Taxation? Asymmetric Tax Effects on Outbound FDI, April 2008.

No. 58 Eicher, T.S: and T. Strobel, Germany's Continued Productivity Slump: An Industry Analysis, March 2008.

No. 57 Robinzonov, N. and K. Wohlrabe, Freedom of Choice in Macroeconomic Forecasting: An Illustration with German Industrial Production and Linear Models, March 2008.

No. 56 Grundig, B., Why is the share of women willing to work in East Germany larger than in West Germany? A logit model of extensive labour supply decision, February 2008.

No. 55 Henzel, S., Learning Trend Inflation – Can Signal Extraction Explain Survey Forecasts?, February 2008.

No. 54 Sinn, H.-W., Das grüne Paradoxon: Warum man das Angebot bei der Klimapolitik nicht vergessen darf, Januar 2008.

No. 53 Schwerdt, G. and J. Turunen, Changes in Human Capital: Implications for Productivity Growth in the Euro Area, December 2007.

No. 52 Berlemann, M. und G. Vogt, Kurzfristige Wachstumseffekte von Naturkatastrophen – Eine empirische Analyse der Flutkatastrophe vom August 2002 in Sachsen, November 2007.

No. 51 Huck, S. and G.K. Lünser, Group Reputations – An Experimental Foray, November 2007.

No. 50 Meier, V. and G. Schütz, The Economics of Tracking and Non-Tracking, October 2007.

No. 49 Buettner, T. and A. Ebertz, Quality of Life in the Regions – Results for German Counties, September 2007.

No. 48 Mayr, J. and D. Ulbricht, VAR Model Averaging for Multi-Step Forecasting, August 2007.

No. 47 Becker, S.O. and K. Wohlrabe, Micro Data at the Ifo Institute for Economic Research – The "Ifo Business Survey", Usage and Access, August 2007.

No. 46 Hülsewig, O., J. Mayr and S. Sorbe, Assessing the Forecast Properties of the CESifo World Economic Climate Indicator: Evidence for the Euro Area, May 2007.

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.