Forecast Evaluation of European Commission Survey Indicators

Christian Gayer*

Abstract

This study examines the contribution of several survey indicators published by the European Commission to forecasting overall economic activity in the euro area. It entails a quantitative evaluation of the information content of seven composite indicators with regard to GDP growth. A preliminary analysis looks at the stationarity and correlation properties of the variables. Based on bivariate VAR-models and the notion of forecast improvement, the methodological approach is two-fold: In a first step, the focussed relations are studied from an ex post perspective. Employing standard and individual Granger-causality tests, an initial assessment of the mean predictive content of the indicators is provided. On the basis of impulse response and variance decomposition analyses, some more light can be shed on the temporal component of the interrelations between the variables. A more informative assessment of the leading indicators' forecast enhancing power is based on out of sample predictive performance. In a second step, therefore, an explorative out of sample scenario is investigated. Attention is turned to the validation and differentiation of the ex post-results. Finally, it is examined to what extent the relations that proved reliable in the explorative scenario could have been useful in individual real-time settings.

The study affirms a useful informative content of the indicators in general and reports encouraging individual results based on the forecast exercises. The forecast potential is, however, limited to the short term. Generally, the predictive information proves better exploitable when using the indicators in annual differences. The Economic Sentiment Indicator (ESI) is found to be most useful up to one or two quarters ahead, while the retail confidence indicator does not help to improve GDP growth forecasts. Moreover, the Business Climate Indicator (BCI) does not prove more informative than the industrial confidence indicator. The remarkably good performance of the building confidence indicator, being interpreted with caution due to signs of spuriousness, may point to the need for further research to clarify the nature of the different confidence indicators.

Key Words: Business cycle, Confidence indicators, Forecasting, Forecast evaluation

JEL Classification: C32, E32, E37

* European Commission, Directorate General for Economic and Financial Affairs, Business Surveys Unit; christian.gayer@cec.eu.int

Views expressed in this paper exclusively represent the positions of the author and do not necessarily correspond to those of the European Commission. The paper was presented at the 27th CIRET Conference in Warsaw, 15-17 September 2004. Comments and suggestions by two anonymous referees are gratefully acknowledged.

Résumé

La présente étude examine la contribution de plusieurs indicateurs d'enquête publiés par la Commission européenne à la prévision de l'activité économique globale dans la zone euro, sur la base d'une évaluation quantitative du contenu d'informations de sept indicateurs synthétiques concernant la croissance du PIB. Une première analyse porte sur les propriétés de stationnarité et de corrélation des variables. Fondée sur des modèles VAR à deux variables et sur la notion d'amélioration de la prévision, la démarche méthodologique est double : dans un premier temps, les relations ciblées sont étudiées d'un point de vue ex post. A l'aide de tests de causalité de Granger standards et individuels, il est procédé à une évaluation initiale du contenu prédictif moyen des indicateurs. Sur la base d'analyses de la réponse impulsionnelle et de la décomposition de la variance, la composante temporelle des interrelations entre les variables est mise d'avantage en lumière. Une évaluation plus instructive du pouvoir d'amélioration de la prévision des indicateurs avancés s'appuie sur l'efficacité prédictive hors échantillon. Dans un deuxième temps, par conséquent, un scénario exploratoire hors échantillon est examiné. L'attention est dirigée vers la validation et la différenciation des résultats ex post. Enfin, on cherche à déterminer dans quelle mesure les relations qui se sont révélées fiables dans le scénario exploratoire auraient pu être utiles dans les différents cadres en temps réel.

L'étude conclut à un contenu informatif utile des indicateurs en général et fait état de différents résultats encourageants obtenus à partir des exercices de prévision. Le potentiel de prévision se limite toutefois au court terme. D'une manière générale, les informations prédictives sont mieux exploitables lorsqu'on utilise les indicateurs sans forme de différences annuelles. L'indicateur du climat économique apparaît comme le plus utile jusqu'à un ou deux trimestres à l'avance, tandis que l'indicateur de confiance dans le commerce de détail n'aide pas à améliorer les prévisions de croissance du PIB. De plus, l'indicateur du climat des affaires ne se révèle pas plus informatif que l'indicateur de confiance des industriels. Les remarquables résultats de l'indicateur de confiance dans le secteur de la construction, interprétés avec prudence car ils pourraient être fallacieux, soulignent la nécessité de poursuivre les recherches afin de clarifier la nature des différents indicateurs de confiance.

1 Introduction

Business cycle indicators have been used as a means to analyse and forecast economic developments from the beginning of business cycle research. Naturally, special attention has been directed to leading indicators, due to their potential contribution to business cycle forecasting. After the introduction of a common monetary policy, business cycle indicators for the euro area have become a focus of interest for politicians, central bankers, economic researchers, managers and other economic agents to gauge aggregate European cyclical developments. Here, the qualitative indicators published by the European Commission (EC) in the framework of the Joint Harmonised Programme of Business and Consumer Surveys (BCS) play an important role.[1]

Despite a general acknowledgement of an empirically observable early warning function of certain economic variables, doubt was cast on the reliability of business cycle indicators time and again with reference either to theoretical or empirical considerations, thereby calling the usefulness of leading indicators into question. Whereas a set of theoretical rationales can be put forward to constitute a leading behaviour of certain economic indicators,[2] empirical studies of the use of indicators in euro area wide forecasting have come to ambiguous or contradictory conclusions.[3] Generally, in such indicator-based forecasting exercises, one has to distinguish between the information content inherent in the indicators themselves and the influence of the model used to exploit and link this potential information to the reference variable at hand.

In a purely linear setting of (vector-)autoregressive models and covering the period up to 1998 only, Fritsche and Marklein (2001) analysed a wider range of qualitative, monetary and other quantitative indicators and their potential to improve forecasts of euro area industrial production, both in-sample and out-of-sample. They conclude that among their indicator set, the qualitative EC survey indicators perform best. Especially the composite *Economic Sentiment Indicator* (ESI) fulfils all examined criteria of a reliable leading indicator, showing forecast improvement up to a time period of six months ahead. In general, however, the authors remain sceptic about the usefulness of leading indicator relations.

Employing a linear setting, too, Banerjee, Marcellino and Masten (2003) analyse the performance of a large set of European and US macroeconomic indicators in euro area inflation and GDP forecasting up to 2000Q4. Averaging forecasting performance across one up to four respectively eight quarter horizons, they find no evidence of forecast improvement over autoregressions in the case of the ESI or the Industrial Confidence Indicator (ICI). Moreover, they find no gains from (dynamic) factor based over single indicator based forecasts, providing support in favour of simple models.

---

1 For an overview of the programme, see European Commission (1997).

2 See for example De Leeuw (1991). For a deeper discussion of the business cycle indicator approach and its use in economic forecasting, see e.g. Gayer (2003).

3 For related studies on the use of survey indicators focusing on individual European countries, see e.g. Etter and Graff (2002), referring to Switzerland and Boyadjian (2002), referring to France, both using indicators stemming from the manufacturing industry. Dion (2002) is an example of the use of consumer survey data in forecasting consumer spending in the euro area.

Restricting their study to one-step ahead forecasts and focussing on the six largest euro area countries, Mourougane and Roma (2002) compare simple indicator based models to ARIMA benchmarks, using also formal significance tests. Based on a sample up to 2000Q4, they report that the ICI and particularly the ESI can be useful in forecasting quarterly GDP growth in the short run. Improvements from the use of time-varying over constant parameter forecasting models appear fairly limited.

In a recent study for the European Commission,[4] a group of Spanish researchers systematically analyses the possibility of forecasting some of the main quantitative macro variables of the euro area using the information given by the entire set of individual and composite BCS indicators. Focussing on the results of GDP and IP forecasting, the inclusion of survey indicators in "augmented models" provides useful information to improve upon univariate benchmark forecasts up to twelve months, respectively four quarters ahead. Whereas the indicator-augmented autoregressions are generally the best performers, simple leading indicator models based on principal components do also generally outperform the best univariate benchmark models. Considering various different forecasting methods, the study reveals that in most cases autoregressions are not outperformed by more complicated techniques such as SETAR or Markov switching regime models. This holds for univariate as well as for the indicator-augmented specifications. The reported results refer to out-of-sample performance during the two-year period from 2001 to 2002.

The general finding of good performance of simple AR models relative to more complicated models is also confirmed by the study of Marcellino (2002), who compares linear with time-varying and non-linear univariate techniques. In the same spirit, the simulated out-of-sample results of Marcellino, Stock and Watson (2001) provide little evidence that GDP and IP forecasts from multivariate models are more accurate than those from univariate AR models. The authors do, however, not use qualitative sentiment indicators among their macroeconomic data set. As regards multivariate forecasting techniques, forecasts based on estimated factors appear to be somewhat more accurate than other methods.

Focussing on production expectations and their relation to industrial production in EU member countries, Lemmens, Croux and Dekimpe (2004) recently find in-sample evidence of a clear predictive value of BCS data in forecasting combined EU production series over the period from 1985 to 2002.

Whereas part of the documented ambiguity may be explained by different estimation and forecasting techniques (recursive vs. rolling estimation, iterating one-step ahead forecasts vs. dynamic h-step estimation) and the different samples and reference series employed, the broad range of results concerning the usefulness of survey indicators for the euro area yet remains puzzling and particularly unsatisfactory. On the other hand, what can be read from any of the above cited studies is that the univariate autoregressive model is apparently quite a tough benchmark for short-term forecasting aggregate output series with any kind of multivariate forecasting model, if the sole judgement criterion is to minimize mean forecast errors. Moreover, the comparably good performance of simple models seems to hold also for multivariate, indicator-augmented forecasting approaches.

4 "Forecasting models currently applied to indicators computed on the basis of survey results", Final Report, November 2003, coordinated by M. Artis and J. Surinach.

Against the sketched background, the aim of the present analysis is to systematically explore the usefulness of composite EC survey indicators in forecasting aggregate economic developments in the euro area. As in the cited studies, "usefulness" will be understood as a significant improvement over autoregressive projections at different horizons. Focussing on real GDP growth as the most comprehensive measure of economic activity, an in-sample evaluation of diverse composite confidence indicators will be complemented by looking at different simulated out-of-sample settings. The fact that different types of composite indicators are under analysis, including the factor model based Business Climate Indicator (BCI), will also allow to explore the usefulness of factor analysis as opposed to more straightforward ways of indicator construction for the particular BCS data under focus.

The paper is organised as follows. Section 2 provides more details on the variables under analysis and the results of some preliminary stationarity and correlation analyses. In Section 3 the in-sample results are reported. Following standard pair-wise and individual Granger-causality tests, impulse response and variance decomposition analyses of bivariate VAR models are employed to shed some more light on the temporal component of the interrelations between the variables under examination. Section 4 reports the simulated out-of-sample forecasting exercise: After presenting an explorative setting of differently parameterised VAR models, two different model selection strategies and their impact on the relative performance of the indicator-based VAR models are investigated. Section 5 summarizes and discusses the main findings and offers some suggestions for further extensions.

2 Data

2.1 European Commission confidence and composite indicators

Survey data are taken from the European Commission BCS programme. Apart from the five individual confidence indicators in industry (INDU), services (SERV), retail trade (RETA), construction (BUIL) and among consumers (CONS),[5] the data set comprises the Economic Sentiment Indicator (ESI) and the Business Climate Indicator (BCI). The first five confidence indicators are simple arithmetic averages of two up to four individual balance series referring to the BCS questionnaires in the sectors at hand. The ESI, being a weighted average of the total of 15 standardised components of the last-mentioned five sector confidence indicators, is specially designed to combine the different cyclical signals to trace overall economic activity, as best measured by GDP growth.[6] The BCI on the other hand, like the confidence

5 For details concerning the EU Harmonised Business and Consumer Surveys, see: http://europa.eu.int/comm/economy_finance/indicators/business_consumer_surveys/userguide_en.pdf

6 The ESI was designed in 1985 to summarise attitudes and judgements of a large number of economic actors from diverse sectors concerning the current and future economic situation. Corresponding to this broad scope of the index, GDP was chosen as the reference variable, tracing the movements of the economy as a whole. The ESI is analysed in its recently revised form, thus especially including the services sector. For a methodological overview of the new, streamlined construction scheme, see: http://europa.eu.int/comm/economy_finance/indicators/business_consumer_surveys/methodological_esi_note_052004_en.pdf

indicator in industry (ICI), is thematically confined to the business cycle in industry: It is based on five balance series from the industrial survey. To extract the common cyclical component of these series, cleaned from idiosyncratic noise, a factor analysis is performed. The BCI results as the first common factor, explaining more than 90% of the variance of the five input series.[7]

All of the employed indicators are thus composite indices, combining information from several single indicators. The rationale behind composite indicators is to strengthen the overall cyclical signal and to enhance the indicator's resilience against idiosyncratic shocks that may occasionally dilute the link between single indicators and the underlying real macroeconomic phenomenon.[8] All series refer to the euro area and are seasonally adjusted. Apart from the services confidence indicator, that is only available from April 1995 onwards, the indicators are available since 1985. The employed activity variable, real GDP, however, is officially obtainable for the euro area from 1991Q1 onwards only and is taken from EUROSTAT (seasonally adjusted). To match the monthly frequency of the indicator series to the quarterly GDP figures, averages are taken over the three monthly indicator values of each quarter.[9]

There are two reasons to believe that the BCS indicators may be useful to improve forecasts for the quantitative GDP variable: First, statistical information from the BCS is available approximately two months in advance of the official statistics. Second, the indicators are inherently related to agents' expectations, so that they should be related to future developments of macroeconomic variables.

2.2 Stationarity analysis

The results of stationarity analysis of the variables are reported in Table 1. As expected, real GDP is integrated of order one, warranting its use in growth rates. The annual growth rates ($gdp = \log(GDP / GDP_{-4})$) appear to be stationary. As the opinion balances mirror cyclically varying perceptions of the economic situation, all of the qualitative sentiment indicators ought to be considered as stationary on a priori grounds.[10] The results of the reported ADF-Tests show, however, that for some of the indicators, the null of non-stationarity

---

7 For details concerning the BCI see: http://europa.eu.int/comm/economy_finance/indicators/business_climate/2001/presentation_climate.pdf

8 For a more in-depth discussion of theoretical rationales behind the construction of composite indicators, see e.g. Emerson and Hendry (1996).

9 Referring to the German IFO Business Climate, Hott, Kunkel and Nerb (2004) argue that the calculation of quarterly from monthly values lead to a loss of prognostic properties due to an alleged time delay. While a smoothing effect is evident, there is however no systematic phase effect of the procedure. Nevertheless, an informative lead of one month could be achieved in principle by relying on the last monthly value only.

10 In more depth, economic sentiment should not be influenced by economic growth - as opposed to the economic cycle as the deviations thereof - since economic agents will constantly adapt their "normal" level of confidence to the growing level of "normal" activity, thereby neutralising the trend component. In practice, of course, the differentiation between trend and cycle complicates the matter.

can only be rejected at the ten percent significance level (CONS, RETA) or even higher (BUIL, SERV). The annual differences of all indicators, however, appear to be stationary.[11]

Table 1 **Unit root tests (augmented Dickey-Fuller)**

	Levels			Annual growth rates		
Series	t-value	Specification	*T*	t-value	Specification	*T*
GDP	-0.04	c, 1	50	-2.94**	c, 1-2	45
ESI	-3.76***	c, 1	54	-5.04***	c, 1	50
BCI	-5.19***	c, 1	54	-7.65***	c, 1	50
INDU	-4.71***	c, 1	54	-6.76***	c, 1	50
CONS	-2.89*	c, 1	54	-4.46***	c, 1	50
RETA	-2.80*	c, 1	54	-2.89*	c, 1	50
BUIL	-2.12	c, 1	54	-2.68*	c, 1	50
SERV	-1.43	c, 1	33	-1.77	c, 1	29

Notes: Maximum sample: 1990:1-2003:4. Critical values according to MacKinnon (1991).
*, **,***: significant at the 10, 5, 1% level .

GDP = Gross Domestic Product; *INDU* = Industry; *BUIL* = Construction
ESI = Economic Sentiment Indicator; *CONS* = Consumer; *SERV* = Services
BCI = Business Climate Indicator; *RETA* = Retail trade

A well known problem of standard unit root tests is that level shifts in otherwise stationary series can lead to erroneous non-rejection of the unit root hypothesis. The qualitative, survey-based BCS series are subject to several minor level shifts due to unavoidable changes in sampling, methodology and questionnaire wording in the continuous process of EU-wide harmonisation. Therefore, and since there is no convincing theoretical reason to believe in non-stationary behaviour of the bounded qualitative survey data, the above-quoted test results for BUIL and SERV are not interpreted as evidence for unit roots in these series.

However, in the following, having the ambiguity of the stationarity results in mind, all indicators will be analysed in both levels and annual differences. The rationale lies in the transformation of the reference series: While the indicators in levels clearly exist in their own right, offering tight correlations to the reference variable (see below) that are promising for the upcoming forecasting exercises, the GDP series has a technical lead when being measured in growth rates. The non-trending nature of the indicators, on the other hand, appears to mirror cyclical trend deviations, showing a lagging behaviour of between three and six months compared to annual differences of the series. To create a level playing field in terms of technical leading behaviour, therefore, it may be crucial to transform indicators and the reference series analogously.

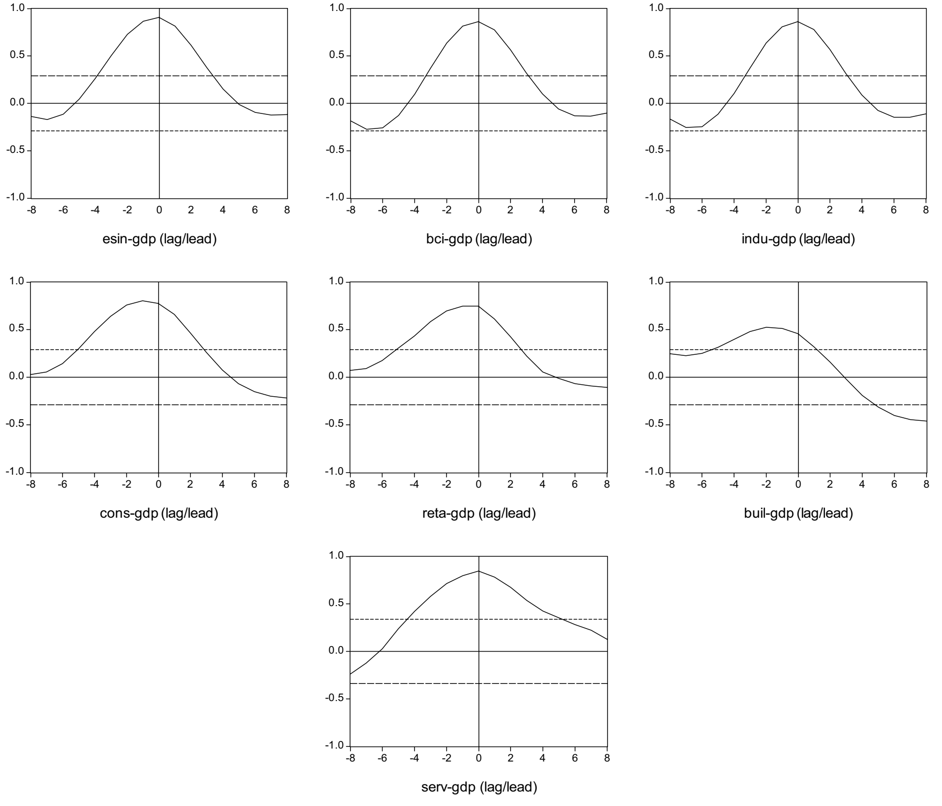
11 An exception is the services confidence indicator, where the null is not rejected. However, here, the test is based on less than 30 observations, which results in very low power of the ADF-test.

2.3 Cross correlation analysis

Figures 1 and 2 report the cross correlations between either untransformed or transformed indicators and GDP growth.

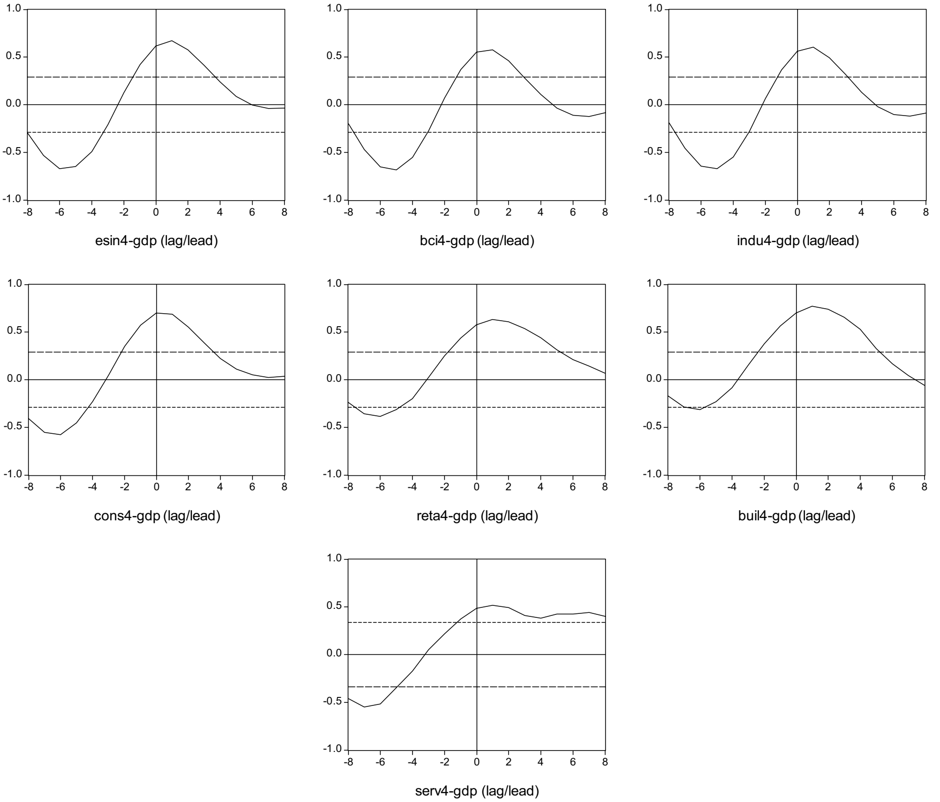
The graphs serve as a first approximation of the lead-lag structures between reference series and indicators, where the maxima of the correlations point to the overall lead or lag of each indicator. The thin lines represent approximate 5% significance bands, with values lying outside these bands indicating significant correlation.

Figure 1 **Indicators in levels: Cross-correlations with GDP growth**



Note: Negative (positive) x-axis corresponds to lagging (leading) behaviour of the indicators with respect to $gdp = \log(GDP / GDP_{-4})$.

Figure 2 **Indicators in annual differences: Cross-correlations with GDP growth**



Note: Negative (positive) x-axis corresponds to lagging (leading) behaviour of the indicators with respect to $gdp = \log(GDP / GDP_{-4})$.

As can be seen, in most cases the correlations indicate coincidence between the level indicator series and GDP growth (ESI, BCI, ICI, SERV), at quite high degrees of correlation between 0.85 and more than 0.90. The correlations are significant up to leads of three quarters. In the case of the consumer and especially the building confidence, there is a clear indication of a lagging behaviour.[12] The retail trade series is at the borderline between coincident and slightly lagging behaviour. As can be expected, when looking at the annually differenced indicator series, their leading properties improve, but mostly at the cost of a lower level of correlation. While the ESI, BCI, ICI and also RETA show leads of one quarter with

12 What should be noted in the case of the building confidence is the significant *negative* correlation at long leads of five and more quarters ("right for the wrong reason").

maximum correlations around 0.60-0.65, the consumer confidence now displays coincidence with a comparably high coefficient of 0.70. The building confidence changes its behaviour in differences most markedly for the better and now displays a lead of one quarter at a distinct maximum correlation of 0.77. The annually differenced services indicator shows a leading pattern, too, but the level of correlation is low.[13]

When interpreting the overall relatively weak signs of leading behaviour of the indicators, even after appropriate differencing, an important fact has to be kept in mind: The presented results are based on indicator and reference series referring to identical time periods (e.g. first quarter 1999), but being available at different points in time. While the indicators for a particular period are available "real time" at the end of that period, the official quantitative data for that period is only published two months later – when two fresh indicator releases are already available. Since the focus of this study is to explore the potential of the qualitative indicators in applied real time settings, based on all available information at the time of forecasting, this informational lead of the indicators has been "internalised" in the analyses of all sections to come. Technically, the indicator series have been shifted backwards by two months in the data organisation, thus mimicking an analogous two months publication lag. Consequently, corresponding indicator and reference values are now simultaneously available with a two months lag.[14] Looking at pairs of GDP and indicator figures does thus no longer discriminate against the indicators in real-time informational terms, as was the case in the conventional comparisons of Figures 1 and 2.

3 In-sample Analysis of Lead-lag Structures

3.1 Granger-causality analysis

To determine whether the inclusion of past values of the indicators would improve the forecast of the reference series in-sample, Granger-causality tests were performed. The tests attempts to determine whether movements in the indicator series precede those in the reference series or vice versa. A common problem in performing Granger tests is the choice of the lag length in the test equations, as the results depend critically on the chosen lag structure. Here, the focus was directed to the correlation properties of the residuals in the test regressions to provide consistent estimation of the (vector)autoregressive coefficients. On the basis of multivariate LM-Tests, five lags were considered necessary in most cases to whiten the residuals. Table 2 reports the results of the standard Granger tests.

13 Due to short data series, the correlations between services confidence indicator series and GDP growth are based on few observations and are thus to be interpreted with greater caution.

14 A one-step ahead forecast will thus be – quite conventionally – a quasi-forecast (now-cast), as it can only be performed two months after the base period for which data is available, taking into account the latest indicator values. Concerning the frequency transformation to quarterly data, the indicators are first shifted by two months and then the quarterly averages are computed.

Table 2 **Standard Granger-tests**

Indicator	Lag length	H_0: Indicator not Granger-causal	H_0: GDP growth not Granger-causal	Result
		Levels		
ESI	5	25.49***	9.75*	↔
BCI	5	17.76***	6.86	→
INDU	5	19.77***	7.74	→
CONS	5	18.15***	10.25*	↔
RETA	5	3.27	6.77	-
BUIL	5	17.92***	11.12**	↔
SERV	5	23.40***	28.53***	↔
		Annual differences		
ESI	5	22.43***	12.47**	↔
BCI	5	22.13***	10.25*	↔
INDU	5	21.32***	11.67**	↔
CONS	5	12.94**	8.61	→
RETA	5	4.44	8.96	-
BUIL	5	8.68	11.61**	←
SERV	4	11.65**	32.36***	↔

Notes: See Table 1 for explanations.
→: Indicator causes GDP, ↔: Feedback, ←: GDP causes Indicator, -: no causation

As can be seen, for all of the indicators but retail confidence there are strong signs of causality from indicators to GDP growth.[15] A result that points to the weaknesses of the approach is the alleged non-causality of the building confidence in annual differences, strongly contrasting to the above reported results of correlation analysis.[16] For many pairs, there are – albeit weaker – signs of reverse causation running from GDP to indicators, too, indicating feedback systems. Though this may cast some doubt on the suitability of the indicators from a business cycle theoretic point of view, the observation that past GDP movements reflect on current indicator values may be helpful in the VAR-based out-of-sample exercises of Section 4, where the indicators will have to be forecasted, too.[17]

15 In the case of the building confidence indicator, the significant test-value is due to the significant *negative* coefficients of long lags.

16 Here, multi-collinearity problems seem to be at play. Individually, the lags 1 to 3 are significantly positive in the test regression.

17 If one believes that only unanticipated shocks can cause changes of the business cycle, only one-way Granger causality relations from indicators to the reference variable are admissible to establish the suitability of an indicator.

The described standard Granger-test suffers from the disadvantage of possible over-parameterisation due to non-significant individual lags of the considered variables. Moreover, it does not allow looking at the significance of individual lags, i.e. the lead structure in more detail. Therefore, individual Granger-tests were performed as follows: First, an autoregression in GDP growth was specified, ensuring white noise residuals and significant coefficients. According to Breusch-Godfrey tests a specification with lags 1, 4 and 5 was sufficient. Then, individual indicator lags from 1 to 4 quarters were added separately and tested for significance by their t-values.[18] Table 3 reports the results for the indicators in levels and in annual differences.

Table 3 **Individual Granger-tests**

Lag	ESI	BCI	INDU	CONS	RETA	BUIL	SERV
			Levels				
1	3.23***	1.13	1.17	0.93	1.21	-0.52	4.45***
2	0.06	-1.31	-1.26	-1.18	0.52	-1.02	2.57**
3	-1.75*	-2.35**	-2.34**	-2.75***	0.70	-1.12	1.41
4	-2.69**	-2.84***	-3.09***	-2.54**	1.35	-1.93*	0.79
			Annual differences				
1	3.33***	2.13**	2.36**	2.89***	-0.14	2.85***	2.39**
2	1.13	0.14	0.23	1.31	0.97	2.92***	1.01
3	0.20	-0.46	-0.51	0.45	0.73	2.49**	0.98
4	0.01	-0.57	-0.70	0.49	1.66	1.49	1.17

Notes: See Table 1 for explanations. Reported are t-values of individual indicator lags. Lags in gdp -autoregression: 1, 4, 5.

Among the relationships with the indicators in levels, only the ESI and the services indicator show significant *positive* contributions to explain future GDP growth – and that only with respect to one or two quarters ahead.[19] For the annually differenced series, all of the indicators but retail confidence significantly impact on GDP growth one quarter ahead. Apart from that, only the building confidence shows further signs of forecast improvement also at the two and three quarter horizon, which contradicts sharply the results of the standard Granger tests but seems more reasonable given the correlation results.

18 In the great majority of cases, white noise residuals were still prevalent after the individual indicator lags were added. Where this was not the case, the Newey/West HAC-covariance matrix correction was used.

19 The significant *negative* entries at lags 3 and 4 for ESI, BCI, INDU and CONS are not to be interpreted as evidence for a negative impact of the indicators on GDP growth at such longer leads. As the correlogram (Figure 1) showed, there are no such significant negative correlations. These negative coefficients rather mirror the dynamic adjustment processes in the autoregressive gdp specification following the initial positive impact of the short-term shock (i.e. the first lag of either the indicators or the highly correlated gdp itself). The significant results especially for lag 4 may also indicate that it is actually the annual differences of the indicators and not the levels that provide relevant information for the gdp process.

The performed Granger-causality tests can be interpreted formally as pointing to the usefulness of the composite indicators for predicting short term GDP growth, at least in differences and excluding the retail trade series. However, it becomes evident that their leading characteristics are confined to shorter horizons. Deducting the internalised two-month publication lead of the BCS series, a significant one quarter-ahead in sample contribution translates into a lead of one month, or rough coincidence in quarterly terms, of the indicators. This broadly reflects the results of the correlation analysis. All in all, the causality analysis indicates a temporally limited potential of the indicators for actual forecasting purposes.

3.2 Impulse response and variance decomposition analysis

To shed some more light on the temporal relations between the indicators and the GDP reference series, and possibly to differentiate the so far quite homogeneous results concerning the individual indicators, two more approaches were followed based on the in-sample relations of the series. First, impulse response analyses of bivariate VAR models, each consisting of GDP growth and one particular indicator, were performed to study the temporal component of the interrelations between the variables. Our main interest is to see how the GDP growth rate reacts in periods $t, t+1, t+2, \ldots$ to an (exogenous) shock in each individual indicator at time t. Second, variance decomposition analysis of the VAR models was employed to examine the relative importance of each individual indicator in explaining the reference series' variability, again as a function of the forecast horizon. Technically, the gdp forecast error variance is decomposed into two components, interpreted as the contributions of (orthogonalised) innovations of gdp itself and those of the individual indicators.[20]

Concerning the parameterisation of the underlying VAR models, the focus has been on uncorrelated residuals as in the Granger-analyses. The selected models are all based on 5 lags. In order to identify the orthogonal structural innovations of the models, the short-term restrictions following from the Choleski-decomposition of the residual covariance matrix were applied.[21] Based on the assumption that all of the indicators display a certain leading character, the order (1) gdp, (2) indicator was imposed on the VAR models, meaning that structural indicator shocks are restricted to impact on gdp movements with a time lag only, while gdp shocks can impact on indicator series immediately.[22]

The results of the analyses are displayed in the annex, both based on indicators in levels and in annual differences. The following conclusions emerge from the impulse response graphs (Figure 3): Generally, significant gdp responses to indicator shocks can be seen up to three quarters ahead, where differenced indicators (column 4) tend to have a somewhat longer lasting impact than levels (column 2). Exceptions are the ESI in annual differences and the services index in levels, where significant gdp reactions are reported up to the four quarter horizon. The building index in annual differences appears to impact at

20 For background and technical details concerning the methods employed in this section, see Lütkepohl (1993). Boyadjian (2002) provides a thorough application of innovation analysis to French business survey data.

21 See for example Lütkepohl (1993).

22 The latter would seem as a necessary precondition for an efficient indicator.

exceptionally high leads between four and six quarters only. In levels, however, like for both retail trade relations, there are no significant reactions of gdp at all. Columns 1 and 3 point to significant responses of gdp to its own innovations generally up to the four quarter horizon. The reported findings are broadly corroborated by the variance decompositions (Figure 4):[23] Generally, the contribution of the indicator innovations to the forecast variance of gdp turns out significantly higher when the annual differences are used (column 3). While the long-turn contribution of indicator shocks to gdp variability stays below that of own gdp innovations for all level indicators but the ESI (column 1), the contribution of differenced indicators is generally much more dominant (especially for ESI, BCI, ICI, but also for the consumer series). In line with the findings of the impulse response analysis, no substantial influence of innovations on gdp forecasts can be seen for the retail trade indicator, neither in levels nor differences.[24] Columns 2 and 4, reporting the variance decompositions for the seven indicators, show that, contrary to GDP growth, the indicators' forecast variability is almost exclusively dominated by own shocks.

While the presented results underline the weak performance of the retail confidence indicator for gdp forecasting purposes, they rarely provide clear evidence for a performance ranking of the other six indicators. If any, the ESI seems to do slightly better than its rivals. However, none of the performed in-sample analyses points to a possible differentiation between the simple average-based ICI and the factor-based BCI. What has been confirmed, on the other hand, is an apparent superior forecast potential of the annual differences of all sentiment indicators compared to their levels.

4 Out-of-Sample Forecasts

4.1 Explorative out-of-sample scenario

A more informative assessment of the leading indicators' forecast enhancing power is based on out-of-sample predictive performance of the underlying heuristic forecasting models. Within a second stage of explorative analysis, therefore, a simulated ex ante-forecasting scenario is investigated. Bivariate VAR models from gdp and individual indicators are first estimated over samples ending in 1999:4 and dynamic forecasts are computed one to four steps ahead.[25] The estimation period is then continuously expanded quarter per quarter until 2003:3. Given that the GDP data ends at 2003:4, a last one-step ahead forecast error can be computed for that period.[26] The exercise produces 16 one-step, 15 two-step,

---

23 The reported results refer to VAR models with indicators coming first and gdp coming second and are qualitatively confirmed in a reversed setting.

24 For the services indicator, the short sample seems to be responsible for the non-stabilising behaviour of the decomposition graphs.

25 The sample start depends on the number of lags in the VAR models and ranges from 92:2 to 94:1.

26 As the latest GDP data is used, the described setting is only an approximation to a real ex-ante scenario that would have to be based on historic, i.e. unrevised data. According to Croushore and Stark (2001), although revisions can lead to considerable differences in forecasting errors in certain periods, there is not necessarily a difference in mean forecasting performance, as measured by RMSEs.

14 three-step and 13 four-step forecast errors. Over these error series, individual h-step root mean square errors (RMSEs) are computed to evaluate the forecasting content of the indicators at the individual horizons. To gauge their usefulness in terms of forecast improvement, the same procedure is run for univariate GDP autoregressions to generate benchmark RMSEs.

Concerning the parameterisation of the models, an explorative setting is analysed first. All of the VAR and AR models will be estimated with one to eight lags, and including a constant. The RMSEs will be presented in the form of modified Theil's U, i.e. relative to those of the benchmark AR model. A value of U less than one indicates an improvement compared to the naïve forecast, whereas values larger than one indicate a worsening of forecast quality. The significance of forecasts improvements is assessed with one-sided Diebold-Mariano (DM) tests.[27]

Table 4 shows the results for the indicators in levels. First of all it appears that, overall, the AR model can only rarely be beaten, even at short horizons where the indicators should provide some additional information. Surprisingly, significant forecast improvements can only be recorded for longer forecasting horizons.

Whereas the fact that the retail confidence indicator does never help to improve over the autoregression is consistent with the in-sample findings, the superior performance of the building confidence apparently contrasts sharply with its comparably poor in-sample performance. However, the observation that past values of building confidence are generally most helpful to improve GDP forecasts is clearly owed to a "right for the wrong reason" phenomenon: As was seen both in the correlation and the individual Granger analysis, there is a strong *negative* correlation of past building confidence with GDP growth. As by means of construction and contents all confidence indicators should be positively related to economic activity (i.e. high confidence index values should indicate a high (future) level of activity), the apparent forecasting content of the confidence level in the building sector is an example of nonsense or spurious correlation, at least with respect to *overall* economic activity.[28]

27 The null of equality of squared forecast errors is tested against the alternative of the indicator-based errors being smaller than AR errors. Following Diebold and Mariano (1995), a truncation lag of h-1 is used for the estimation of the spectrum of the loss-differential $d = e_{\mathrm{AR}}^2 - e_{\mathrm{VAR}}^2$. An advantage of the DM test is that it poses only weak requirements concerning the forecast errors: they may be non-normal, non mean zero and serially and contemporaneously correlated. Since it is an asymptotic test, of course, in the present setting of less than 20 observations, the results must be interpreted with caution.

28 Theoretically, one may argue that confidence in construction might indeed be a good indicator of activity in the construction sector, but that the latter either behaves counter-cyclically with respect to overall activity or might follow a more or less idiosyncratic cyclical path. The latter may be rationalised by the fact that total investment in construction aggregates such different activities as housing, non-residential building and public civil engineering, each driven by individual cyclical (or other, including discretionary policy-related counter- or non-cyclical) dynamics.

Table 4 **RMSE of AR and VAR models with indicators in levels**

Horizon	Lags	RMSE	RMSEs relative to AR models							
(h)	(p)	AR	AR	ESI	BCI	INDU	CONS	RETA	BUIL	SERV
1	1	0.45	1	1.24	1.28	1.36	1.23	1.03	1.05	**0.94**
1	2	0.44	1	1.01	**0.92**	**1.00**	1.04	1.03	1.06	**0.98**
1	3	0.43	1	1.13	1.13	1.17	1.16	1.07	1.11	1.27
1	4	0.46	1	1.41	1.37	1.45	1.39	**1.00**	1.06	1.54
1	5	0.46	1	1.06	1.06	1.09	***0.84***	1.19	**0.95**	1.54
1	6	0.43	1	1.08	**0.93**	1.05	**0.94**	1.13	1.02	2.39
1	7	0.43	1	1.36	1.01	1.19	1.34	1.13	1.05	-
1	8	0.45	1	1.21	1.18	1.15	1.37	1.17	1.20	-
2	1	0.70	1	1.28	1.38	1.43	1.34	1.08	1.10	1.02
2	2	0.68	1	**1.00**	**0.94**	**0.99**	1.12	1.08	1.06	1.20
2	3	0.66	1	1.12	1.12	1.11	1.33	1.11	1.09	1.16
2	4	0.75	1	1.19	1.16	1.18	1.59	1.02	1.02	1.44
2	5	0.83	1	1.10	**0.94**	**0.94**	**0.86**	1.21	***0.85****	1.71
2	6	0.79	1	1.19	**0.89**	**0.92**	**0.88**	1.16	**0.87**	2.23
2	7	0.77	1	1.57	**0.93**	1.13	1.23	1.19	**0.87**	-
2	8	0.81	1	1.37	1.19	*1.13*	1.34	1.40	1.26	-
3	1	0.88	1	1.35	1.42	1.45	1.47	1.10	1.18	1.10
3	2	0.86	1	1.10	1.01	1.07	1.30	1.10	**0.92****	1.62
3	3	0.88	1	1.15	1.07	1.05	1.55	1.08	1.03	1.55
3	4	0.96	1	1.29	1.08	1.13	1.82	1.06	**1.00**	1.18
3	5	1.17	1	1.16	**0.97**	**0.95**	**0.97**	1.27	**0.82***	1.66
3	6	1.15	1	1.24	**0.87**	**0.87**	**0.96**	1.18	***0.80****	2.05
3	7	1.10	1	1.61	1.01	1.16	1.30	1.25	**0.83**	-
3	8	1.21	1	1.52	1.30	1.19	1.38	1.37	1.27	-
4	1	1.07	1	1.37	1.38	1.40	1.52	1.10	1.22	1.16
4	2	1.07	1	1.15	1.03	1.06	1.40	1.10	**0.90*****	1.92
4	3	1.07	1	1.18	1.06	*1.01*	1.65	1.06	*1.01*	1.87
4	4	1.18	1	1.38	1.19	1.21	1.94	1.13	1.01	1.28
4	5	1.52	1	1.21	**0.99**	**0.96**	1.02	1.28	**0.80***	1.28
4	6	1.50	1	1.27	**0.87***	**0.87***	**0.96**	1.18	**0.77***	1.91
4	7	1.44	1	1.62	1.03	1.18	1.23	1.25	***0.70*****	-
4	8	1.58	1	1.55	1.40	1.29	1.29	1.27	1.20	-

Notes: See Table 1 for abbreviations.
Relative root mean square errors (RMSEs) below 1 (unrounded) are marked bold.
The smallest value at each horizon is in italics.

*, **,*** indicates significant forecast improvement at the 10, 5, 1% level (DM-test).
No VAR models for SERV at 7 and 8 lags due to data restrictions.

Table 5 **RMSE of AR and VAR models with indicators in annual differences**

Horizon	Lags	RMSE	RMSEs relative to AR models							
(h)	(p)	AR	AR	ESI	BCI	INDU	CONS	RETA	BUIL	SERV
1	1	0.45	1	**0.85**	**0.90**	**0.88**	**0.81***	1.18	**0.96**	1.16
1	2	0.44	1	**0.73****	**0.80***	**0.78***	**0.73****	1.10	**0.99**	1.10
1	3	0.43	1	**0.82**	***0.76***	**0.83**	**0.79***	1.14	1.03	1.75
1	4	0.46	1	1.06	**0.99**	1.08	1.02	1.14	1.03	1.91
1	5	0.46	1	***0.72****	**0.74***	**0.79**	**0.88**	1.16	**0.92***	-
1	6	0.43	1	**0.82**	**0.90**	**0.92**	1.07	1.17	**1.00**	-
1	7	0.43	1	**0.96**	**0.93**	1.01	1.19	1.06	**0.86***	-
1	8	0.45	1	**0.97**	1.25	1.22	1.08	1.08	**0.81***	-
2	1	0.70	1	**0.92**	**0.99**	**0.98**	**0.85**	1.18	**0.86**	1.31
2	2	0.68	1	**0.86**	**0.93**	**0.88**	***0.80****	1.14	**0.97**	1.35
2	3	0.66	1	**0.97**	**0.91**	**0.98**	**0.87**	1.18	1.01	2.03
2	4	0.75	1	**0.83**	**0.81**	**0.83**	**0.84**	1.14	**0.90**	2.28
2	5	0.83	1	**0.82**	**0.90**	**0.89**	**0.91**	1.22	**0.86****	-
2	6	0.79	1	**0.89**	**0.99**	**0.94**	1.01	1.30	**0.97**	-
2	7	0.77	1	1.20	1.05	1.06	1.16	1.32	**0.93**	-
2	8	0.81	1	1.17	1.28	1.17	1.04	1.34	**0.94**	-
3	1	0.88	1	**0.93**	1.05	1.05	**0.82**	1.04	***0.71*****	1.44
3	2	0.86	1	1.05	1.15	1.08	**0.91**	**1.00**	**0.82***	1.63
3	3	0.88	1	1.13	1.13	1.16	**0.96**	1.03	**0.92**	2.33
3	4	0.96	1	**1.00**	**0.98**	1.02	**0.90**	1.09	**0.88**	2.72
3	5	1.17	1	**1.00**	1.05	1.03	1.01	1.26	**0.81*****	-
3	6	1.15	1	1.09	1.13	1.07	1.08	1.35	**0.91**	-
3	7	1.10	1	1.40	1.18	1.16	1.16	1.41	**0.89*****	-
3	8	1.21	1	1.36	1.31	1.21	1.12	1.44	**0.92**	-
4	1	1.07	1	**0.91**	1.03	1.03	**0.83**	**0.94****	***0.69****	1.46
4	2	1.07	1	1.15	1.21	1.15	1.04	**0.92**	**0.81****	1.71
4	3	1.07	1	1.26	1.22	1.24	1.11	**0.97**	**0.89***	2.23
4	4	1.18	1	1.11	1.07	1.12	**0.97**	1.01	**0.85**	2.90
4	5	1.52	1	1.05	1.09	1.10	1.05	1.22	**0.79*****	-
4	6	1.50	1	1.15	1.22	1.17	1.10	1.31	**0.91**	-
4	7	1.44	1	1.16	1.19	1.20	1.14	1.43	**0.87*****	-
4	8	1.58	1	1.21	1.23	1.20	1.05	1.53	**0.91**	-

Notes: See Table 4 for explanations.

Table 5 displays the results of the forecast comparisons based on annual differences of the confidence indicators. It is evident, that all of the indicators but the retail and services confidence do now markedly improve the GDP forecasts at the one and two quarter horizon. While, generally, at those horizons a slight advantage in forecast improvement can be observed for the ESI,[29] it is again the building confidence – now in differences –, that clearly outperforms both the AR model and the other indicators at the longer horizons of three and four quarters. This finding is in line with the individual Granger-tests, where only for the differenced building index significant positive contributions were documented, and the impulse response graphs. The weaker performance of the services index in the out-of-sample scenario seems to be owed to the shorter estimation sample; the first forecast calculations are based on estimation samples of only three to four years.

Considering the performance of the BCI, which is based on a factor analysis of the three series making up the industrial confidence plus two other industrial survey series, there is some evidence of improved forecast content when compared to the industry confidence indicator in levels. After transformation into annual differences, however, the simple confidence index is slightly more often superior to the extracted business climate factor. Altogether, there does not seem to be much gain from cleaning the industrial survey series from idiosyncratic noise by common factor extraction from the angle of possible (GDP) forecasting power improvement.[30]

Comparing the BCI to the ESI, the annual changes of the ESI are found to provide the best forecasting contribution in the majority of cases. As could be expected due to its broader coverage, for forecasting movements of *overall* economic activity, the ESI seems to be better suited than the industry-related BCI. Looking at the two industry-related indicators (BCI, ICI) on the one hand and the consumer confidence index on the other, an overall similar information content with respect to GDP growth emerges.

Lastly, comparing the two forecasting scenarios, the annually differenced indicators do clearly provide more forecast-relevant information than the original level series do. This seems to be attributable, partly, to the improved leading characteristics that are a technical consequence of the taking of annual differences. Even then, however, there is generally no exploitable information beyond the two quarter horizon. The most remarkable exception from this rule is the changes of the building confidence that, somewhat surprisingly, help to significantly improve GDP forecasts even at the three and four quarter horizons. Considering the implausible or at least difficult to interpret results using the confidence levels, some caution seems to be warranted as to the robustness of this finding.

---

29 The ESI-based model shows the lowest RMSE among the indicator models in five out of the 16 cases (h=1,2; p=1,2,…,8) and also the lowest average over these 16 RMSEs.

30 This result is, of course, specific to the present analysis and does not necessarily carry over to more general approaches of GDP forecasting. The benefits of factor extraction are likely to become more visible when based on larger and more diverse input data sets. For applications of factor models in the area of BCS data based forecasting, see e.g. Bruno and Malgarini (2002) and De Mulder and Dresse (2002).

4.2 Two real-time forecasting scenarios

The results presented so far provided an overview of the general and mean forecast relevant information content of the BCS indicators. The explorative setting did, however, circumvent the critical question of how to best exploit the indicators in real time forecasting, i.e. what model to choose for generating the short-term forecasts. Focussing on the results of the VAR models with indicators in annual changes, there is no obvious result of what VAR models perform best across the different forecasting horizons; at longer horizons (3 and 4 quarters), parsimonious models seem to perform better. Though backed partly by intuition and the results of previous studies, however, this information would not have been available in real time forecasting situations. To select benchmark and indicator models based on information that was readily available at the individual points of time during the forecasting sample, the Schwarz Information Criterion (SIC) was employed first. At each estimation step, therefore, that model was chosen as a basis for forecasting that produced the lowest SIC then, where a maximum lag of eight was considered.

Table 6 shows the results: In the simulated ex ante exercise, the ESI produces the smallest mean forecasting error at the one and two quarter horizons, where the first reduction over the AR is significant at the 10% level. Apart from retail trade and services, improvements over autoregressions could also have been realised with the other indicators. At the longer horizons, again, the building confidence indicator seems to perform best, albeit not significantly better than the benchmark model.

Looking behind the presented results, in the large majority of cases, the information criterion selected (vector)autoregressive models with a quite high number of five lags. The comparison with other models from Table 5 reveals that in terms of the RMSE, the VAR(5) and especially the AR(5) models do generally perform relatively poorly out of sample. Apparently, the in-sample fit, as mirrored by the SIC, is not very informative with regard to the out of sample forecasting performance.[31]

Therefore, in a second step, a different procedure for model selection was employed, based on past out of sample performance of candidate models with one up to eight lags: The forecasting process is initialised using SIC-selected models covering the first estimation sample (ending in 1999:4 as before). Mimicking a real time scenario, as soon as the first h-step forecast errors of the selected and the other seven alternative models become available (i.e. h quarters after the initial forecasts), the model with the lowest absolute error is used for producing the next h-step forecast. For the one step-ahead forecast, for example, only the

[31] Generally, such a result is not very encouraging as to the stability of the employed heuristic models. However, as was reported in the introduction, a trade-off between in sample fit and forecasting performance of competing forecasting models is one of the universal findings of many studies dealing with forecast evaluation of indicator based models, where often the simplest models produced the best out of sample projections. Moreover, here, it seems to be special characteristics of the reference series that influence the outcome of the SIC selection. The observation that the fifth lag improves the in sample fit markedly, especially in the AR but also most VAR models, points to possible distortions from the use of seasonally adjusted GDP raw data.

first error to enter the RMSE calculation stems from the SIC-selected model; the second and all subsequent errors are based on those models that performed best in the prior (i.e. first, second, etc.) quarter. For the four step-ahead forecasts, on the other hand, its takes one full year before it can be gauged what model produced the lowest initial error. The first four entries of the error record are thus filled with errors from SIC based specifications.

Table 6 **RMSE of SIC selected models (indicators in annual differences)**

Horizon	RMSE	RMSEs relative to AR							
(h)	AR	AR	ESI	BCI	INDU	CONS	RETA	BUIL	SERV[#]
1	0.46	1	***0.72****	**0.74***	**0.79**	**0.88**	1.16	**0.94**	1.43
2	0.83	1	***0.82***	**0.90**	**0.89**	**0.91**	1.22	**1.00**	1.51
3	1.17	1	**1.00**	1.05	1.03	1.01	1.26	***0.96***	1.64
4	1.52	1	1.05	1.09	1.10	1.05	1.22	***0.91***	1.64

Notes: See Table 4 for explanations. [#] Services confidence indicator in levels

It follows from the quasi-real time set up of the procedure that the selection process lags h quarters behind the optimum, i.e. RMSE-minimising selection. A potential advantage of this refined selection approach is that forecasts at different horizons are produced based on horizon-specific information on past forecast performance.[32] Especially in the case of the four-quarter ahead forecasts, on the other hand, the long time delay of one year may obstruct this potential advantage. Table 7 presents the results of the approach, where the retail and services indicators were excluded due to their previous weaker performances.

Comparing the column giving the absolute RMSE of the AR model with that of Table 6 shows that the described model selection procedure clearly improves over the SIC, i.e. in sample-related, selection mechanism.[33] However, it also becomes apparent that the improvement seems to be less pronounced in the case of the indicator augmented models, as the relevant error ratios mostly increase when compared to Table 6. The forecast enhancing power of the indicators is thus less visible in the present setting. At the short horizon, only the ESI produces visible, albeit not significant, error reductions over the AR model. As before, it is the building confidence that performs best at the longer horizons, even significantly outperforming the AR benchmark at the four quarter horizon. To sum up, the results of the refined simulated ex-ante forecasts broadly corroborate the previous findings.

32 It followed from the explorative results that forecasts at longer leads tend to be better based on parsimonious models, while this was not generally the case for the shorter horizons. The results of the presented selection approach confirm this finding: Across all models, the average chosen lag decreases as the forecast horizon increases.

33 According to the DM-test, the reported improvements of the AR forecasts are significant at all forecasts horizons (5% level).

Table 7 **RMSE based on past out of sample performance (indicators in annual differences)**

Horizon	RMSE	RMSEs relative to AR					
(h)	AR	AR	ESI	BCI	INDU	CONS	BUIL
1	0.38	1	***0.89***	1.18	1.17	**0.98**	1.05
2	0.63	1	1.16	1.24	***0.95***	1.09	1.11
3	0.96	1	**1.00**	1.34	1.26	**0.89**	***0.83***
4	1.17	1	1.03	1.24	1.36	1.12	***0.86****

Notes: See Table 4 for explanations.

5 Discussion and Conclusions

This paper assessed the usefulness of the composite indicators of the European Commission based on the Business and Consumer Surveys programme in short-term GDP growth forecasting. The main finding is that the overall forecasting potential of the survey indicators is generally restricted to the one and, less evidently, two quarter horizon. Considering the publication lag of GDP figures, this translates into usefulness in now-cast, respectively one quarter ahead forecast scenarios. These findings broadly mediate previous results by other studies ranging from reported practical uselessness to forecast improvements up to the four quarter horizon.

Looking at individual indicators, the ESI, being the broadest index in sectoral coverage and as could therefore be expected, seems to have most potential for improving forecasts of GDP growth. The retail confidence indicator, on the other hand, is the only indicator that can not contribute to forecast improvements. In the case of the services confidence indicator, presumably related to estimation problems due to shorter data availability, encouraging results could only be found in the in-sample exercises. Discriminating between the simple average-based industrial confidence indicator and the factor model-based BCI, no visible enhancement of forecasting properties emerged from the elimination of idiosyncratic noise due to the factor approach. Furthermore, the consumer confidence indicator proved likewise informative regarding future GDP growth as the two industry-related series.

Unexpected results were encountered for the confidence in the construction sector: While a significant *negative* correlation at long leads was responsible for (only spurious) GDP forecast improvements when the index was used in levels, the annual changes of the building index consistently contributed to forecast improvements even at the three and four quarter horizons, thereby clearly outperforming the other indicators. Since there is hardly convincing a priori reason to believe that exclusively changes in the building sentiment should lead economic activity by three and more quarters, some doubt is warranted as regards the robustness of these particular findings. Further research seems necessary to explore the characteristics and the meaning of the results from the building survey in more detail. In this broader context, some more light should be shed on the general question of whether the composite survey indicators themselves measure levels or changes of confidence and, relatedly, what kind of reference series the extracted confidence variables refer to, both from a theoretical and empirical perspective.

In practical terms, it appeared that the use of annual changes of indicators was mostly preferable to the use of levels in terms of forecast improvement. Even for those indicators where the correlation at the one or two quarter lead was lower after differencing, the short-term forecasting properties mostly improved (ESI, BCI, ICI).

The most important caveat of the analysis is the relatively small sample, especially in relation to the number of parameters in the less parsimonious models of the explorative setting. Moreover, special events as e.g. the creation of EMU or the 11 September 2001 terror attacks, potentially deserving a thorough econometric modelling within the employed estimation and forecasting set-up, have not been taken into account. The comparison of simple AR with indicator-augmented VAR models is used as a heuristic statistical instrument to gauge potential forecasting gains, certainly not as an attempt to generate valid forecasting models.

Possible extensions of the study could be the inclusion of single indicators from the BCS programme as well as the use of industrial production as a second reference series in a monthly setting. Going further into that direction, individual sector-related reference series could be focused to look more specifically at the prognostic content of each indicator, e.g. retail confidence in relation to household consumption. Moreover, while the focus was on mean relationships between variables over the whole sample, the present work could usefully be complemented by looking at individual indicators' behaviour at cyclical turning points in particular.

References

Banerjee, A.; M. Marcellino and I. Masten. 2003. "Leading Indicators for Euro-Area Inflation and GDP Growth," CEPR Discussion Paper No. 3893, May.

Boyadjian, S. 2002. "VAR modelling of Demand Components and Business Surveys," Paper Presented at the 26th CIRET Conference, Taipei, October.

Bruno, G. and M. Malgarini. 2002. "An Indicator of Economic Sentiment for the Italian Economy," Paper Presented at the 26th CIRET Conference, Taipei, October.

Croushore, D. and T. Stark. 2001. "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics* 105, pp. 111-30.

De Leeuw, F. 1991. "Toward a Theory of Leading Indicators," in: *Leading Economic Indicators: New Approaches and Forecasting Records*, Lahiri, K. and G.H. Moore eds. Cambridge, pp. 15-56.

De Mulder, J. and L. Dresse. 2002. "The Construction of Leading and/or Coincident Indicators for Belgian Economic Activity," Paper Presented at the 26th CIRET Conference, Taipei, October.

Diebold, F.X. and R.S. Mariano. 1995. "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13:3, pp. 253-63.

Dion, D.P. 2002. "Does Consumer Confidence Forecast Household Spending? – The Euro Area Case," Paper Presented at the 26th CIRET Conference, Taipei, October.

Emerson, R.A. and D.F. Hendry. 1996. "An Evaluation of Forecasting using Leading Indicators," *Journal of Forecasting* 15:4, pp. 271-91.

Etter, R. and M. Graff. 2002. "Estimating and Forecasting Production and Orders in Manufacturing Industry from Business Survey Data: Evidence from Switzerland, 1990-2000," ETH Zurich, KOF Working Paper No. 66, August.

European Commission. 1997. "The Joint Harmonized EU Programme of Business and Consumer Surveys," European Economy, Reports and Studies No. 6.

Fritsche, U. and F. Marklein. 2001. "Leading Indicators of Euroland Business Cycles," DIW Discussion Paper No. 238, Berlin, January.

Gayer, C. 2003. Prognoseevaluation konjunktureller Gesamtindikatoren für Deutschland: Eine ökonometrische Analyse von F.A.Z.-Konjunkturindikator, Handelsblatt-Frühindikator und Ifo-Geschäftsklima-Index. Frankfurt/Main.

Hott, C.; A. Kunkel and G. Nerb. 2004. "Turning Point Predictions with the Ifo Business Climate," in: *Composite Indicators of Business Activity for Macroeconomic Analysis,* Matkowski, Z., ed. Warzaw, pp. 167-84.

Lemmens, A.; C. Croux and M.G. Dekimpe. 2004. "On the Predictive Content of Production Surveys: A Pan-European Study," K.U.Leuven Research Paper, January.

Lütkepohl, H. 1993. *Introduction to Multiple Time Series Analysis*, 2nd ed., Berlin et al.

MacKinnon, J.G. 1991. "Critical Values for Cointegration Tests," in: *Long-Run Economic Relationships,* Engle, R.F. and C.W.J. Granger, eds., Oxford.

Marcellino, M. 2002. "Forecasting EMU Macroeconomic Variables," CEPR Discussion Paper No. 3529, October.

Marcellino, M.; J.H. Stock and M.W. Watson. 2001. "Macroeconomic Forecasting in the Euro Area: Country Specific versus Area-Wide Information," IGIER-Università Bocconi Working Paper No. 201, October.

Mourogane, A. and R. Moreno. 2002. "Can Confidence Indicators be Useful to Predict Short Term Real GDP Growth?", ECB Working Paper No. 133, March.

Annex

Figure 3 **Impulse Response Analysis**

Indicators in levels | Indicators in annual differences

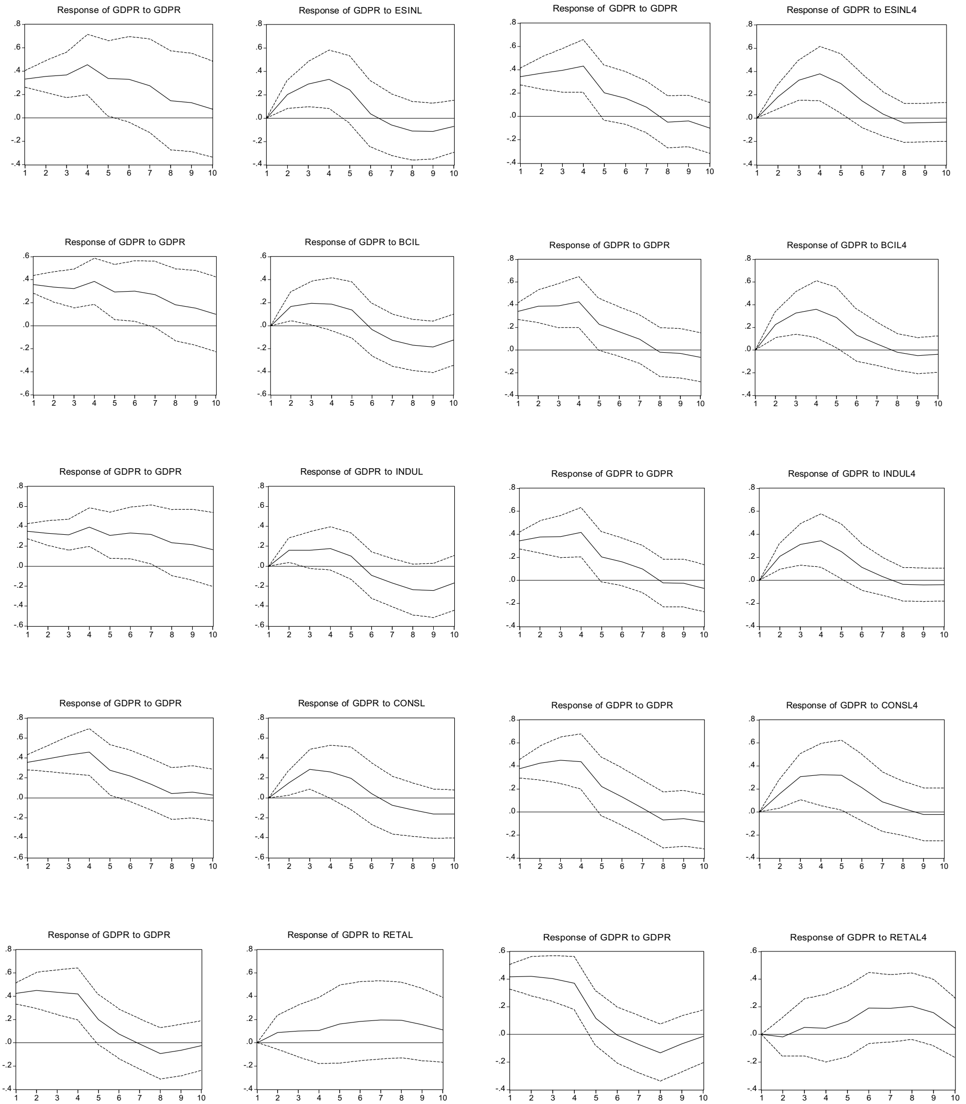
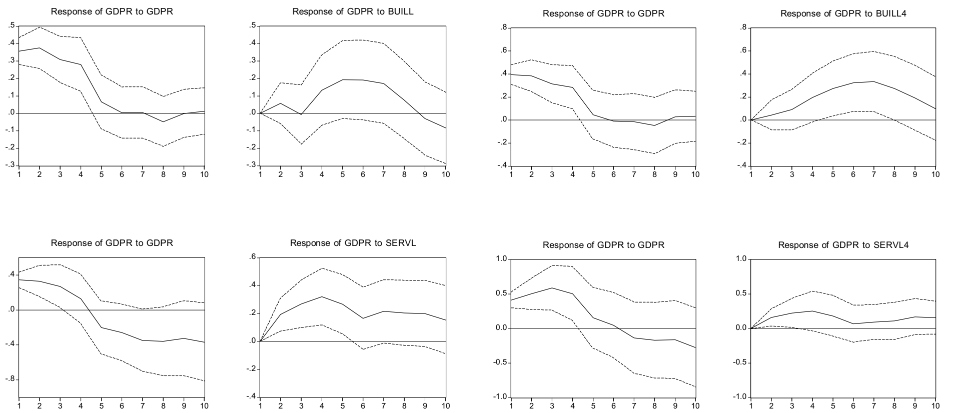


Figure 3 **Impulse Response Analysis (continued)**

Indicators in levels | Indicators in annual differences



Notes: Responses of current and future values of GDP growth (denoted GDPR in the figures) to one unit orthogonal shocks. Columns 1 and 2 refer to VAR models of GDPR and individual indicators in levels, columns 3 and 4 to VAR models with indicators in annual differences

Figure 4 **Variance Decompositions**

Indicators in levels

Indicators in annual differences

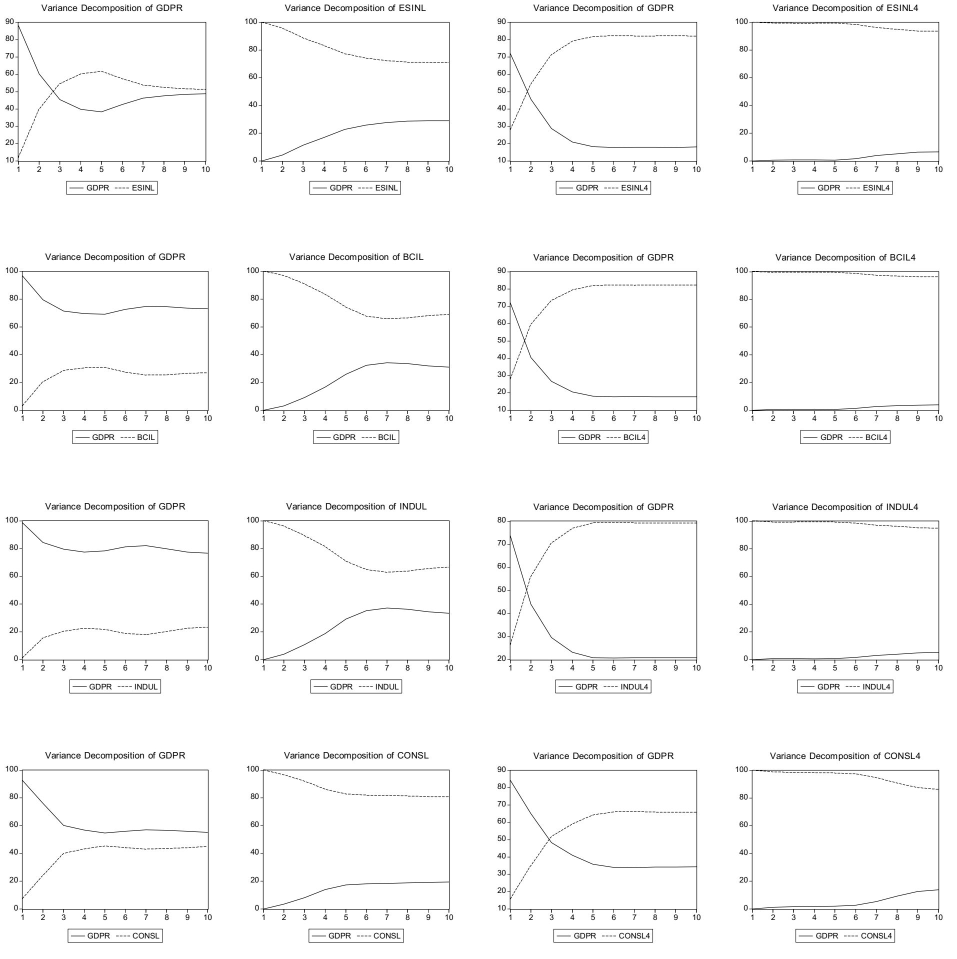
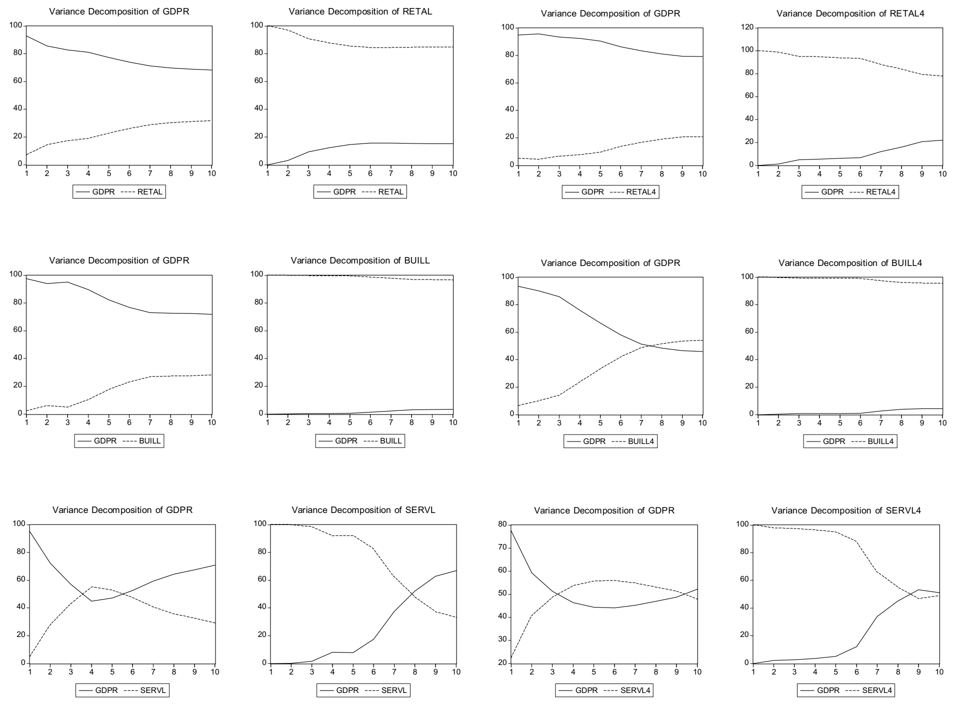


Figure 4 **Variance Decompositions (continued)**

Indicators in levels Indicators in annual differences



Notes: Attribution of the forecast variance to orthogonalised innovations of GDP growth (denoted GDPR in the figures) and indicators for increasing forecast horizons. Columns 1 and 2: bivariate VAR models using indicators in levels, columns 3 and 4: models using indicators in annual differences

Contents

From:

Journal of Business Cycle Measurement and Analysis

Access the journal at:

https://doi.org/10.1787/17293626

Please cite this article as:

Gayer, Christian (2006), "Forecast Evaluation of European Commission Survey Indicators", *Journal of Business Cycle Measurement and Analysis*, Vol. 2005/2.

DOI: https://doi.org/10.1787/jbcma-v2005-art2-en