An experimental sentiment indicator for the euro area – the relevance of broad-based sectoral shifts in economic sentiment[1]

By Andreas Reuter[2]

Abstract

The Economic Sentiment Indicator (ESI), published by the European Commission on a monthly basis, is a powerful tool for tracking year on year GDP growth. However, its performance is weaker when GDP growth is expressed in quarter on quarter changes. This paper investigates whether cross-sector survey data gathered in the framework of the harmonised EU Business and Consumer Survey (EU BCS) can be combined in a way different from the ESI construction method, with the explicit aim of boosting correlation with q-o-q GDP growth. The construction method under investigation deviates from the ESI in that i) only the survey questions best correlated with q-o-q GDP growth are used and ii) the q-o-q change of the selected questions is amplified through multiplication with a constant in case a critical amount of the selected questions (at least 8 out of 11) changes in the same direction as the average. The logic of the latter is that changes in the average of the survey questions should be taken more "seriously" (i.e. amplified) in case they are broad-based, i.e. reflected in many questions (and sectors) underlying the indicator. The resulting experimental indicator indeed achieves promising results: Its coincident and leading correlation with q-o-q GDP growth improves significantly compared to the ESI.

In a subsequent step, the paper examines the merits of the experimental sentiment indicator for the purpose of nowcasting q-o-q GDP growth in the euro area at the end of the third month of a given quarter. Departing from an autoregressive benchmark model, the performance of a bi-variate model containing only the experimental indicator is tested. Subsequently, slightly more complex models, which add some "hard" data as predictor variables, are presented. The hard data include macro-economic variables (unemployment, etc.) as well as sector-specific variables (index of industrial production, etc.) which are sufficiently timely available at euro area level. Using a top-down testing approach, the non-significant variables are removed to produce parsimonious models. All models, including the simple bi-variate one, are shown to outperform the benchmark model in terms of MAE, RMSE and different hit-ratios. The differences are most pronounced from 2008q2 onwards (i.e. from the onset of the financial crisis). The best-performing of the new models combines the sentiment indicator with industrial production growth and a measure of volatility on the European stock markets.

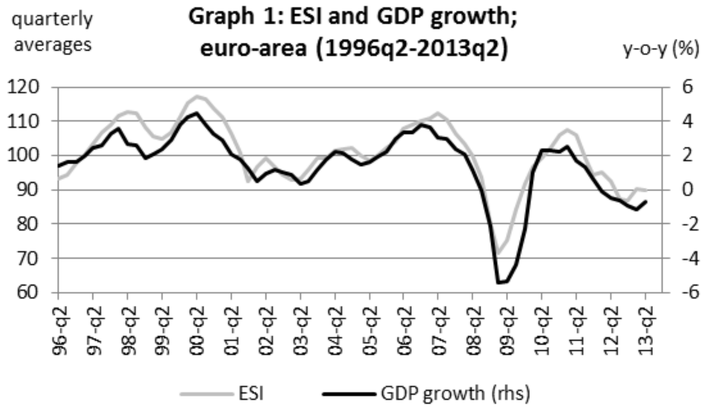
[1] This paper is based on the highlight section of DG ECFIN's European Business Cycle Indicators (4th quarter 2012). Large parts are literal quotes from that publication. The analysis has been developed in cooperation with Marian Neagu (Romanian National Forecast Commission).

[2] European Commission, DG ECFIN, Economic Analyst in 'Business and Consumer Surveys and Short-term Forecast'

I. Background

The principal aim of the Joint Harmonised EU Programme of Business and Consumer Surveys (BCS) is to provide political decision-makers, economists and economic agents with information about the current state of the economy. The survey results also regularly feed into models that aim to now- and forecast economic growth in the short-term. The relevance of BCS data is mainly due to their timeliness compared to official statistics (so-called hard data), which are usually published with significant delays. The EU BCS programme covers five sectors of the economy (industry, services, construction, retail trade and consumers) and dedicates to each of them a specific "confidence indicator", which summarises sector-specific tendencies in a single number. To provide a summary measure of confidence throughout the entire economy, all survey questions used for the construction of the sector-specific confidence indicators are aggregated into a single measure – the Economic Sentiment Indicator (ESI). The ESI is an excellent tool for tracking year-on-year GDP growth in the euro area.[3]

Graph 1: ESI and GDP growth; euro-area (1996q2-2013q2)



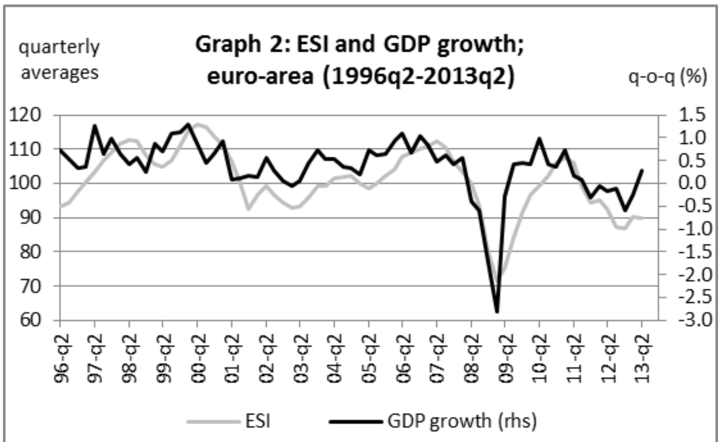
Note: monthly BCS data are converted into quarterly by averaging the balances over 3 months. GDP figures refer to y-o-y changes (%).
Source: European Commission.

As is readily apparent from Graph 1, the ESI evolves smoothly and mimics the up- and downswings of GDP growth with high precision, which is also reflected in a coincident correlation of 0.93 over 1996q2 to 2013q2. However, the ESI's performance is comparatively weaker with respect to quarter-on-quarter GDP growth.[4] This observation holds in particular in comparison with Markit Economic's Purchasing Managers' Index (PMI)[5]. Graphs 2 and 3 display, respectively, the ESI and the PMI in comparison with quarter-on-quarter euro-area GDP growth.

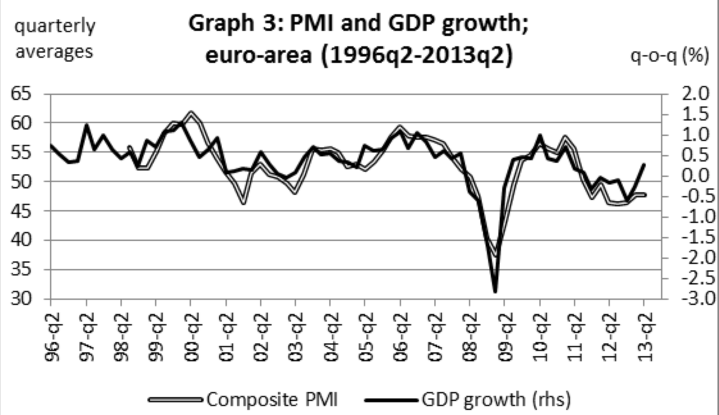
[3] See highlight section of European Business Cycle Indicators, 3rd quarter 2012.

[4] See e.g. The UniCredit Research Toolbox of 13 June 2012

[5] Throughout the analysis, Markit Economic's Final Eurozone Composite Output Index has been used.



Note: monthly BCS data are converted into quarterly by averaging the balances over 3 months. GDP figures refer to q-o-q changes (%). Source: European Commission.



Note: monthly PMI data are converted into quarterly by averaging the balances over 3 months. GDP figures refer to q-o-q changes (%). Source: European Commission, Markit Economics.

The better tracking performance of the PMI with respect to quarter-on-quarter growth is among others revealed by its behaviour at cyclical turning points: For example, the ESI signals the downturn starting in 2007q1 with a two-quarter delay, whereas the PMI signals a drop only with a one-quarter delay. By the same token, the PMI mimics the sharpness of the upswing starting in 2009q2 almost perfectly, while the ESI underestimates its pace.

When making such comparisons, it should be borne in mind that tracking q-o-q GDP growth is not an explicit aim of the ESI and its construction has not been trimmed to produce particularly high correlations with that reference series. Furthermore, the q-o-q and y-o-y GDP growth series obviously differ in terms of their volatility and amplitude, so that a time-series very highly correlated with y-o-y growth can hardly perform equally well with respect to q-o-q growth[6].

Against this background, this paper presents a different way of combining EU BCS data such that the resulting indicator allows a better tracking of quarter on quarter developments in euro-area GDP. The proposed indicator is constructed as a quarterly measure, i.e. quarterly averages of the underlying monthly balance series of survey questions are used for its computation.

In a subsequent step, the paper examines the merits of the experimental sentiment indicator for the purpose of nowcasting q-o-q GDP growth in the euro area at the end of the third month of a given quarter. Starting from a simple model which uses only the experimental indicator as predictor variable, more complex models are developed which add some hard data to the model.

II. The construction of an experimental sentiment indicator

Given that the reference series to be tracked (EA q-o-q GDP growth) is a wide measure encompassing overall economic activity, the construction of the new experimental indicator sticks to the fundamental principles on which the ESI is based: i) the indicator shall be an average of several EU BCS questions, ii) the questions shall stem from all five sectors surveyed, iii) each sector shall be allocated a weight, broadly reflecting the relative importance of the economic sector in GDP as well as the degree to which the sectoral questions are correlated with GDP growth. In order to be able to attribute possible improvements in the tracking of q-o-q GDP growth to the modifications detailed below, the weighting scheme is unchanged compared to the ESI[7].

The major changes in calculating the new indicator compared to the ESI are the following: i) the selection of survey questions underlying the indicator is tuned to track q-o-q developments in GDP (pre-selection) and ii) under certain conditions, the quarterly changes in the indicator are amplified through multiplication with a constant, the logic of which will be explained below.

II. 1. Choice of the right set of survey questions

The procedure for choosing the questions to be included in the indicator follows a "best-performer" approach. All EU BCS questions, aggregated at euro area level, are correlated with q-o-q growth in i) GDP and ii) the respective sectoral reference series[8]. Subsequently,

[6] Note that the PMI has a lower coincident correlation with y-o-y GDP growth than the ESI (0.85 vs. 0.94 for the period 1998q3 to 2013q2).

[7] Industry is weighted with 40%, services with 30%, consumers with 20%, construction with 5% and retail trade with 5%.

[8] For industry: Gross Value Added (GVA) in manufacturing; for services: GVA in services, for construction: GVA in construction; for consumers and retail trade: household and non-profit institutions serving households (NPISH) final consumption expenditure.

for each sector, three new confidence indicators are constructed: A first one is based on the two questions with the highest correlation with the reference series/GDP, a second one on the three questions best correlated with the reference series/GDP[9] and a third one based on all forward-looking questions of the sector. The logic of the latter is that most end-users of the EU BCS programme can be assumed to attach particular importance to the leading properties of an indicator, which are likely to be enhanced through inclusion of forward-looking questions. In a subsequent step, for each sector, the confidence indicator yielding the highest correlation with the sectoral reference series and GDP is selected. The questions making up these confidence indicators are the ones to be used for the computation of the new indicator. Compared to the ESI, questions change in every sector. Industry, retail trade and consumers see, respectively, two of the questions currently used for the ESI discarded and a new question added. In case of both services and construction, one of the questions currently used fails the selection procedure and an additional one is added to construction, while none to services. The questions finally selected are the following:

- production expectations in industry
- past production in industry *
- past demand in services
- expected demand in services
- consumers' expected financial position
- consumers' expected general economic situation
- consumers' expected level of major purchases *
- expected sales in retail trade
- expected orders placed with suppliers in retail trade *
- current order books in construction
- past building activity in construction *

Compared to the ESI, four of the questions, marked with an asterisk, are new. Combining the results (balance statistics) of these eleven questions in one indicator[10], results in slight improvements in tracking q-o-q GDP growth compared to the ESI (Table 1).

[9] Note that the ranking of questions is identical, no matter if correlation with the reference series or with q-o-q GDP growth is used as a ranking criterion.

[10] At this stage of the analysis, the construction method does not differ from the ESI, except that a different set of survey questions is used and quarterly averages of the questions' balance series are used as input.

Table 1: Correlations of ESI and new indicator with GDP growth (q-o-q)

	ESI	ESI based on alternative questions	improvement
Coincident correlation	0.72	0.77	7%
Leading correlation	0.49	0.54	10%
Note: correlation coefficients are computed over the period 1996q2 – 2013q2; coincident correlation is computed using current quarter values for both survey and hard data while for the leading correlation hard data is shifted one quarter ahead.			

II. 2. Amplifying changes reflected by many survey questions

To understand the rationale of the second step of the indicator construction, some theoretical explanations are warranted. The motivation for step 2 of the construction rests on the following considerations: i) there is an infinite number of possible combinations of q-o-q changes in the survey questions making up the indicator which all result in the same q-o-q change of the composite indicator; ii) if the composite indicator reports a positive q-o-q change, this can be the result of all underlying survey questions moving up, of just one survey question moving up (obviously, sharply to outweigh the losses in the other questions), or of a few questions moving up and a few ones moving down; iii) the extent to which a given positive q-o-q change in the composite indicator translates into GDP growth can be assumed to be higher if the increase in the indicator is reflected by moderate increases in a large amount of underlying survey questions (and sectors) rather than significant increases, which are confined to just a few survey questions. Obviously, the above considerations hold analogously for downward shifts in the composite indicator.

These considerations suggest that changes in the composite indicator should be taken more "seriously", when reflected by many underlying survey questions. Practically, it is suggested that the q-o-q changes of the composite indicator (positive or negative) are multiplied by a constant larger 1 (i.e. amplified), if a critical amount of questions changes in the same direction as the composite indicator does. As regards the critical amount of questions, the threshold should be chosen such that it is sufficiently restrictive (e.g. it would be conceptually hard to defend that if 6 out of 11 questions move in the same direction as the indicator, the movement is so broad-based that amplification is justified). At the same time, the value should be low enough to trigger amplification of the cyclical signal in a sufficient number of quarters. Based on these criteria, a threshold of 8 questions has been chosen, leading to amplification in 62% of cases in the period 1996q2 to 2013q2 [11]. The value by which the changes in the indicator are multiplied is set at 3, being the integer which maximises the correlation of the new indicator with q-o-q GDP growth over the analysed sample (1996q2 to 2013q2).

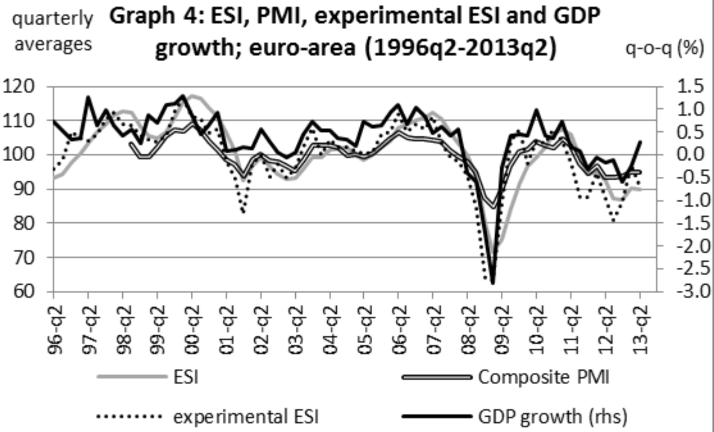
[11] If the threshold was set at 9, 10 or 11 questions, it would be triggered in 49%, 35% or 16% of cases.

Based on these parameter settings, the new indicator can be constructed. In a first step, for each quarter, the standardised weighted questions are summed up. In a second step, the q-o-q change of this aggregate measure is calculated. Subsequently, a "trigger variable" is calculated, which takes the value 1 if the respective quarter reports 8 or more survey questions changing in the same direction as the aggregate compared to the previous quarter. In a fourth step, the q-o-q change of the sum of standardised weighted questions is multiplied by 3 if the trigger variable for the respective quarter is 1; otherwise it is left un-altered. The respective new q-o-q change of quarter t is then added to the original (i.e. un-altered) sum of standardised weighted questions of quarter t-1. The result is a new composite indicator, which either simply takes the value of the sum of the 11 standardised weighted questions, in case the respective quarter saw less than 8 questions moving in the same direction as the aggregate, or is artificially increased / decreased otherwise. For presentational ease, the new indicator is standardised and re-scaled to have a long-term average of 100 and a standard deviation of 10 (as the ESI).

III. The tracking performance of the new indicator

Graph 4 presents the new experimental indicator (indicated by a dotted line) along with the ESI, the PMI and the reference series, i.e. q-o-q GDP growth. Compared to the ESI, the new indicator is characterised by a higher q-o-q volatility, which in most instances is more in line with the reference series. A first example is the temporary recovery of GDP growth during its downswing from 2000q2 to 2001q2, which is not captured by the ESI, but well-reflected in the new indicator. Other examples include the downturn starting in 2007q1, which both the ESI and the new indicator report only in the third quarter of 2007.

Graph 4: ESI, PMI, experimental ESI and GDP growth; euro-area (1996q2-2013q2)



Note: monthly BCS data and PMI (rescaled to mean 100) are converted into quarterly by averaging the balances over 3 months. GDP figures refer to q-o-q changes (%). Source: European Commission.

The new indicator, however, drops significantly sharper, which is more in line with the behaviour of GDP after the intensification of its drop from 2008q2 onwards. Finally, the new indicator perfectly mimics the pace (i.e. slope) of the recovery starting in 2009q2, while the ESI underestimates its intensity. When comparing the new indicator's performance to the PMI, the added value is less evident, since the PMI is already highly correlated with q-o-q growth. Nevertheless, there are some instances, where the new indicator outperforms the PMI. One example is the see-sawing of GDP (up-down-up-down) in the period 2006q1 to 2007q2. The PMI just reports an increase, followed by a downturn, after which the series follows a horizontal path. The new indicator, by contrast, perfectly mimics the see-sawing movement. The downswings of 2010q3 and 2011q2 are further illustrations of the new indicator's superiour performance – notably of its leading properties. In fact, both downswings are signalled by the new indicator two and one quarters before they actually materialise. The PMI just achieves a coincident reporting of these downturns.

The assessment can be formalised by taking a look at Table 2, which reports the coincident and leading correlations of the PMI and the new indicator with q-o-q GDP growth, as well as the improvements of the new indicator compared to the PMI in terms of percentage increase in correlation. To make the analysis more robust, the observation period is split into several sub-periods. It turns out that the new indicator is on the same level as the PMI in terms of coincident correlation. To put these results into perspective, it should be recalled that the ESI achieves a coincident correlation of 0.75 with q-o-q GDP growth over the period 1998q3 to 2013q2. When correlating GDP growth of quarter t with the respective indicator of quarter t-1 (leading correlation), the new indicator clearly outperforms the PMI.

Table 2: Correlations (leading correlations) of Composite PMI and experimental ESI with GDP growth (q-o-q)

time-period	PMI	exp. ESI	improvement
98q3-02q1	0.73 (0.54)	0.75 (0.52)	3% (-4%)
02q2-07q1	0.86 (0.63)	0.86 (0.69)	0% (10%)
07q2– 13q2	0.87 (0.59)	0.89 (0.72)	2% (22%)
98q3– 07q1	0.78 (0.58)	0.79 (0.58)	1% (0%)
98q3– 13q2	0.86 (0.65)	0.89 (0.75)	3% (15%)

Note: coincident correlation is computed using current quarter values for both survey and hard data while for the leading correlation the hard data is shifted one quarter ahead.

IV. The forecasting performance of the new indicator

In the following, the merits of the experimental ESI for the purpose of nowcasting euro area q-o-q GDP growth are tested. Given that the experimental indicator is constructed on basis of the quarterly averages of BCS data and the corresponding quarter on quarter changes, the indicator is only available at the end of the last month of a given quarter. Accordingly, also any nowcast of GDP relying on the indicator cannot be conducted earlier than at the end of the quarter to be nowcast. While earlier nowcasts would be desirable, it should be borne in mind that GDP figures are released some 45 days after the end of the reference quarter so that the merits of the nowcasts presented below in terms of information lead are still substantial.

IV. 1. Model selection based on regression analyses

To assess the performance of the experimental ESI in nowcasting GDP, an appropriate benchmark model is needed. Starting from an autoregressive model with 3 lags, elimination of non-significant lags results in a benchmark model where euro area q-o-q GDP is explained by one lag of GDP (see Table 3, models 1 to 3). That model achieves an R-squared of 0.48 for the period 1999q2 to 2013q2[12].

In a next step, q-o-q GDP growth is explained by a constant and the experimental indicator. In line with its high correlation with GDP growth (see previous chapter), the experimental ESI proves highly significant and produces an R-squared of 0.80 (see model 4 in Table 3), thus clearly outperforming the autoregressive benchmark model. The indicator with its relatively wide sectoral coverage thus achieves to capture a large share of the variation in GDP[13].

Based on those reassuring results, it is subsequently tested whether the performance can be enhanced when adding certain hard data to the model. Given that the model can only be run at the end of the third month of a given quarter, various hard data like industrial production, retail sales and the unemployment rate of the quarter's first month are already released and can thus be included in the model. The hard data whose value-added for the bridge model has been investigated are i) the index of industrial production (IPI), ii) the index of production in construction (IPC), iii) the index of turnover in retail trade, except of motor vehicles and motorcycles (reta), iv) the value of euro area exports (extra- and intra-euro area), v) the unemployment rate (unem), vi) new passenger car registrations (car reg) and vii) the VStoxx as a proxy of the volatility on the European stock market[14]. All variables, except for the unemployment rate and the VStoxx, are expressed as q-o-q changes.

Including all of the above-described hard data in the regression (see Model 5) and subsequently excluding all variables that are non-significant (top-down approach), results in a specification where q-o-q GDP growth is explained by the experimental sentiment indicator, q-o-q industrial production growth, q-o-q construction production growth and the VStoxx in lagged levels (see Model 6)[15]. The R-squared is at 0.91.

---

[12] The historic estimation period has been restricted to achieve comparability across the models, which differ in terms of the time-periods for which their respective explanatory variables are available (VStoxx, for example, is only available from 1999q1 onwards).

[13] The R-squared of a model where q-o-q GDP growth is explained by a constant and the classical ESI (period 1999q1 to 2013q2) achieves an R-squared of 0.57, which is in line with the comparatively weak correlation of the ESI with q-o-q GDP growth diagnosed in a previous section of the article.

[14] The VStoxx is the square root of the implied variance of Euro Stoxx 50 real-time options prices.

[15] The VStoxx only proved to be significant when lagged by one quarter, reflecting that it is derived from the variance in options (i.e. forward-looking investment vehicles).

With a view to producing parsimonious models,[16] it is investigated in a subsequent step whether the R-squared remains broadly unchanged if one of the variables (except for the sentiment indicator) is dropped. The results (see Models 7-9) suggest the following conclusions: i) more parsimonious models without a much lower R-squared are possible, ii) they can be achieved by either dropping VStoxx (-1) or q-o-q growth in construction production, iii) q-o-q growth in industrial production should, however, remain part of the regression, since dropping it results in a noticeably lower R-squared[17].

Table 3: regression results

dependent variable: GDP(q-o-q)		Model 1	2	3	4	5	6	7	8	9	10
Explanatory variables	const	nsig	nsig	nsig	sig	nsig	sig	sig	sig	sig	sig
	GDP(-1)	sig	sig	sig							
	GDP(-2)	nsig	nsig								
	GDP(-3)	nsig									
	exp ESI				sig	sig	sig	sig	sig	sig	sig
	ΔIPI					sig	sig		sig	sig	sig
	ΔIPC					sig*	sig	sig	sig		
	VStoxx (-1)					sig	sig	sig		sig	
	Δreta					nsig					
	Δexp intra/extra EA					nsig					
	Unem					nsig					
	Δcar reg					nsig					
Adjusted R-squared		0.49	0.49	0.48	0.80	0.93	0.91	0.84	0.90	0.90	0.88

blank: not included – nsig: not significant – sig: significant (at the 5% level)
*significant at 5.2% level
Sample period:

- model 1: 1999q4 to 2013q2
- model 2: 1999q3 to 2013q2
- model 3: 1999q2 to 2013q2
- models 4, 8, 10: 1999q1 to 2013q2
- model 5: 2000q2 to 2013q2
- models 6, 7, 9: 1999q2 to 2013q2

Based on the last observation, it is also tested whether both IPC and lagged VStoxx can be dropped at the same time (see Model 10). The adjusted R-squared drops by 0.02 points,

[16] Parsimony is important for a precise estimation of the model's parameters, enhances its interpretability and robustness and reduces the scope for data mining, i.e. the tailoring of a model to optimise its fit to historical data rather than its out-of-sample forecasting performance.

[17] This was to be expected given that the correlation of q-o-q growth in industrial production with q-o-q GDP growth (for the period 1999q1 to 2013q2) is at 0.89, while q-o-q growth in construction production and the lagged VStoxx correlate less well with q-o-q GDP growth (0.32 and -0.45 respectively).

suggesting the model would be inferiour. However, given the small magnitude of the decrease in the R-squared, as well as the aim of parsimony and, last but not least, keeping in mind that in-sample and out-of-sample performances can differ, the model is not (yet) discarded.

To sum up, the most promising bridge models are models 8, 9 and 10. The next section will examine their out-of-sample nowcasting performance, taking the AR(1) model as a benchmark. Finally, to be able to single out the added value of the hard data, nowcasts will also be reported for model 4, which takes the experimental ESI as only predictor variable.

IV. 2. The out-of-sample properties of selected models

As is well known from the forecast literature, a model performing well with regard to in-sample criteria (e.g. adjusted R-squared) does not necessarily do the same in an out-of-sample context. To shed light on the out-of-sample properties, historical nowcasts for the benchmark models and the most promising bridge models are conducted in a simulated real-time scenario. The sample used to estimate the parameters of the models has a static starting point (1999q2) and a dynamic end point, which is always the quarter preceding the quarter to be nowcast. Nowcasts are conducted for every quarter of the period 2004q1 to 2013q2.

To simulate the nowcasting behaviour under realistic conditions, for a given nowcast of quarter q, only the data available up to (including) the last month of quarter q are considered[18]. While this does not restrict the use of EU BCS soft data, which is published at the end of each month, it does restrict the extent of hard data that can be used for the nowcast, since they are usually released with a significant time-lag. Concretely, at the end of a given quarter, only IPI and IPC of the first month of the quarter are available.

As an answer to this complication, the growth rates of the industrial production and construction production indices are calculated as the percentage change of the first month of quarter q (for which the indices' values are available) compared to the readings of the first month of quarter q-1. The method produces satisfactory proxies of the actual q-o-q growth rates, as shown by high correlations of the proxy series with the quarterly series[19]. The measure of volatility on the European stock markets is available at the end of the reference quarter[20] so that no proxies for it must be constructed.

---

[18] It should be noted that the entire simulation is based on hard data downloaded at the end of August 2013, which means the data is revised compared to the original data available in the last month of the respective nowcast quarter. Thus, the analysis is essentially a pseudo out-of-sample exercise.

[19] In the period 1999q1 to 2013q2, the correlations between the proxy series and the actual q-o-q growth series are 0.91 for the industrial production index and 0.69 for the construction production index. It should be noted that also a different specification has been tested, namely growth rates calculated as the percentage increase of the reading of month 1 of quarter q compared to the average of the monthly readings of quarter q-1. Using the resulting growth rates to produce nowcasts results only in minor changes compared to the nowcasts presented here.

[20] The VStoxx is an intra-day measure. To construct a measure of quarterly volatility, the daily average readings of the VStoxx have been averaged over the entire quarter.

The performance of the bridge models is judged based on four criteria: First of all, the root mean squared error (RMSE) and the mean absolute error (MAE). The difference between the two measures is that the former "punishes" for comparatively large forecast errors. Intuitively, the RMSE might be understood as a measure not only considering the MAE, but also the dispersion of the errors. Furthermore, the analysis is complemented by two types of hit-ratios. The first is the percentage of nowcasts that predict the correct sign of GDP growth (which is important for determining whether the economy is in technical recession), the second is the percentage of nowcasts that correctly identify an acceleration/deceleration in GDP growth.

A look at the left section of Table 4 allows an assessment of the nowcasting performance of the auto-regressive benchmark model. The MAE over the entire nowcasting period (grey cells) is at 0.39, meaning that, on average, the nowcasts are about 0.40 percentage points off the actual GDP growth rate. Also the RMSE (0.67) is rather high. It goes without saying that the level of the MAE and RMSE differs depending on the nowcasting period investigated. Both measures are smallest in the pre-crisis period of 2004q1 to 2008q1, while highest in the turbulent crisis period of 2008q2 to 2009q4.

With this benchmark in mind, one can turn to the right-side of the table, which sheds light on the nowcasting performance of the experimental-ESI-based models. It turns out that the simple model, which uses the experimental ESI as only predictor variable, clearly outperforms the autoregressive model. The MAE is driven down by roughly a third, to 0.27, while the RMSE, which imposes a higher "punishment" on large forecast errors, is almost halved (to 0.36)[21]. A look into the scores for the sub-periods provided in the table (pre-crisis, crisis, post-crisis), shows that the improvement comes almost entirely from more precise nowcasts in the crisis-period. In fact, both the MAE and the RMSE for the crisis-period more than halved compared to the benchmark model. The seeming merits of the experimental ESI in turbulent times are rather intuitive, recalling that it amplifies the average change in its constituent survey questions in case at least 8 of the 11 survey questions move in the same direction. Thus, the indicator is explicitly designed to capture "momentum", and there were two cases of acute "momentum" in the crisis period 2008q2 to 2009q4 (notably the sharp acceleration of the GDP contraction in 2008q4 and the sudden deceleration of the contraction in 2009q2).

Models 8, 9 and 10 are an attempt to increase the nowcast precision even more through the addition of hard data to the experimental ESI model. It turns out that the hard data, indeed, adds information on top of the already well-performing experimental ESI model. The best models in terms of MAE (model 9) and RMSE (model 10) respectively achieve reductions in the two quality measures of some 25% over the entire nowcast-sample (grey cells). A view on the sub-periods shows that the improvements stem mainly from a significantly better performance in the post-crisis period (depending on the model, reductions in the MAE/RMSE

[21] If the classical ESI is used instead of the experimental ESI, the MAE and RMSE over the entire period 2004q1 to 2013q2 are at 0.38 and 0.54 respectively. The experimental ESI has thus superiour out-of-sample properties, while the nowcasting performance of the classical ESI does not differ much from a simple AR(1) model.

of 46% and 35%) and, only to a lesser extent, in the crisis-period. This is in line with the above considerations according to which the construction method of the experimental ESI makes it particularly effective in crisis times and, thus, the added value of hard data will be more pronounced in economically calmer periods.

Table 4: Out-of-sample properties of the models

	Benchmark model	New models			
number (from Table 1)	3	4	8	9	10
specification	GDP (-1)	exp. ESI	exp. ESI, IPI,IPC	exp. ESI, IPI,VStoxx(-1)	exp. ESI, IPI
MAE	0.39	0.27	0.21	0.20	0.21
2004q1-2008q1	0.18	0.16	0.15	0.16	0.15
2008q2-2009q4	1.10	0.49	0.42	0.41	0.42
2010q1-2013q2	0.28	0.28	0.16	0.15	0.17
RMSE	0.67	0.36	0.29	0.27	0.26
2004q1-2008q1	0.22	0.19	0.20	0.20	0.18
2008q2-2009q4	1.45	0.61	0.51	0.43	0.44
2010q1-2013q2	0.34	0.34	0.22	0.23	0.23
Hit ratio 1	32/38	33/38	35/38	36/38	34/38
2004q1-2008q1	17/17	17/17	17/17	17/17	17/17
2008q2-2009q4	5/7	6/7	7/7	6/7	6/7
2010q1-2013q2	10/14	10/14	11/14	13/14	11/14
Hit ratio 2	14/37	26/37	28/37	30/37	29/37
2004q2-2008q1	5/16	11/16	9/16	11/16	11/16
2008q2-2009q4	5/7	7/7	7/7	7/7	6/7
2010q1-2013q2	4/14	8/14	12/14	12/14	12/14
The figures are based on nowcasts over the interval 2004q1 to 2013q2.					

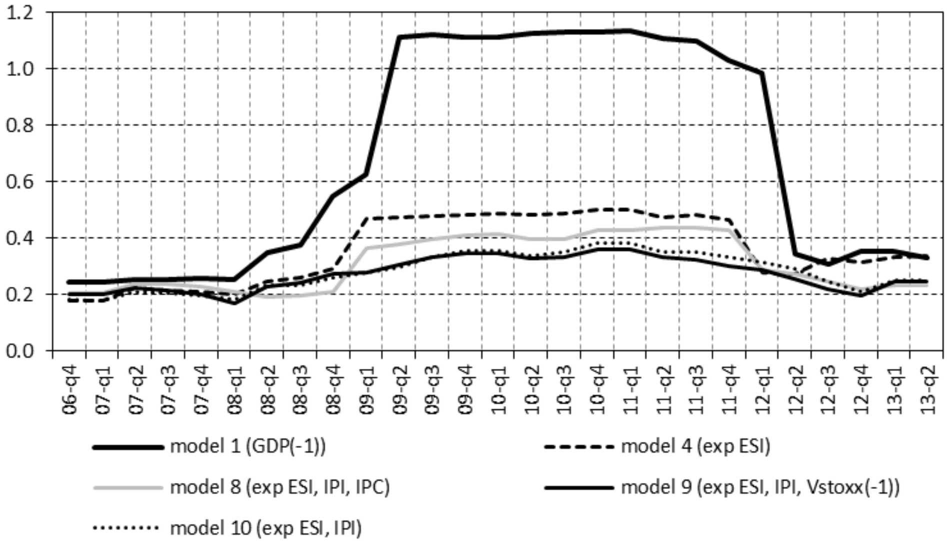
Turning to the amount of nowcasts that identify the right sign of GDP growth (hit ratio 1), the experimental ESI model (model 4) achieves a score of 33 out of 38. The addition of hard data results in improvements of up to 3 points (model 9). In case of the amount of nowcasts correctly determining an acceleration/deceleration in GDP growth (hit ratio 2), a score of 26 out of 37 nowcasts is driven up to 30 out of 37 (model 9). In line with the analysis regarding MAEs/RMSEs, the improvements both in absolute and relative terms are highest in the post-crisis period.

Having diagnosed i) a superiority of model 4 (only experimental ESI as predictor) vis-à-vis the auto-regressive model and ii) the possibility of enhancing the nowcasting performance, in particular for the post-crisis period, even further through the addition of a selection of hard data, the next question to answer is whether any of the models using hard data is clearly better than the others.

Being the weakest of the models using hard data in terms of both RMSE and hit ratio 2, model 8 can hardly be argued to be the best of the models. Models 9 and 10 are comparable in terms of their MAEs/RMSEs (differences of 0.01). However, model 9 achieves to correctly identify the correct sign of GDP growth in two more cases and an acceleration/deceleration in economic activity in one more case. The relative advantage of model 9 is in particular visible in the recent past: over the period 2010q1 to 2013q2, the model nowcasts the wrong sign of GDP growth only in one case, while model 8 and 10 mis-predict the sign in 3 cases. It is worth pointing out that identification of the right sign of GDP growth was particularly challenging in the period 2010q1 to 2013q2, where GDP growth hovered around 0. These observations advocate in favour of including a lagged operator of the VStoxx index (the only difference between models 9 and 10) in the model. Thus, model 9 is the preferred one.

Graph 5 gives a good visual summary of the analysis of the different models' nowcasting performance. It reports, for each model and each quarter starting in 2006q4, the RMSEs based on the nowcasting errors of the preceding 3 years.

Graph 5: Rolling RMSEs by model

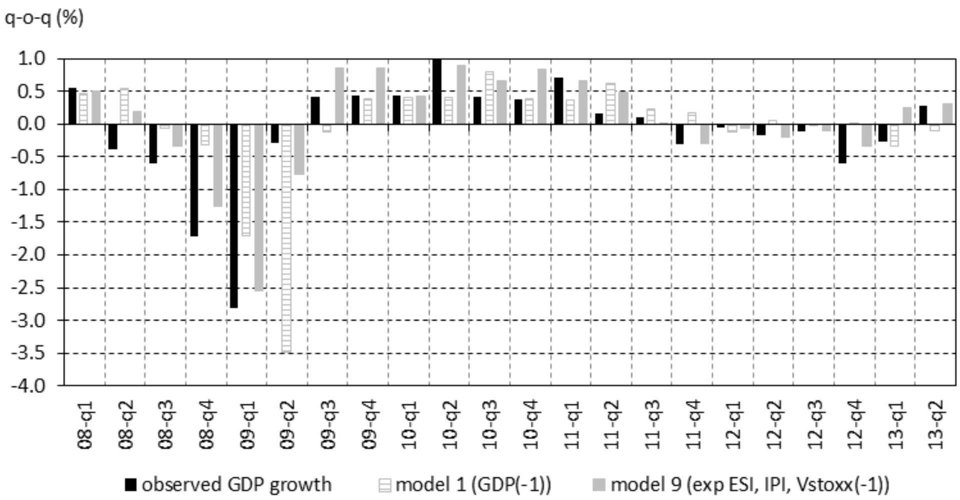


Note: The graph reports for each model and each quarter the RMSE based on the nowcasting errors of the preceding 11 quarters and the current quarter.

The graph shows that the autoregressive model (model 1) has seen nowcasting errors ballooning when the economic crisis hit (three-year rolling RMSE rising from 2008q2 onwards). Model 4, using the experimental ESI only, constitutes a significant leap forward compared to the benchmark model. However, the comparative advantage has faded over the last quarters, resulting from the calming economic conditions, in which the particular strength of the experimental ESI in detecting "momentum" is less of an asset. All models adding hard data to the experimental ESI achieve improvements, albeit smaller ones. Model 9, which adds IPI growth and lagged VStoxx to the experimental ESI, is persistently the best model from 2010q3 to 2012q4.

The observations can be summed up in Graph 6, which reports actual GDP growth, as well as nowcasts produced by the auto-regressive model (the worst model) and model 9 (the best one). The merits of model 9 become particularly obvious when focussing on the last quarters (2011q4 to 2013q2), where GDP growth was hovering around 0 and, accordingly, nowcasting the correct sign of GDP growth posed a particular challenge. In fact, model 9 missed the right sign of GDP growth just once and almost perfectly mimicked the magnitude of the growth rates. This is in sharp contrast to the performance of the auto-regressive model.

Graph 6: Nowcasts of GDP growth produced by bridge models; euro area (2008q1-2013q2)



Note: The nowcasts are produced by a pseudo out-of-sample exercise using soft- and hard data downloaded on 30/08/13. The in-sample period for the nowcasts is 1999q1 to (incl.) the quarter preceding the quarter to be nowcast. Source: European Commission.

V. Conclusion

The point of departure for this paper was that the European Commission's Economic Sentiment Indicator is an almost perfect measure for tracking year-on-year GDP growth in the euro area. However, when GDP is expressed in quarter-on-quarter growth, the performance has been shown to be weaker – especially in comparison to the PMI. An attempt has therefore been made to test whether EU BCS data can be exploited in a different way with the explicit objective of achieving a good tracking of q-o-q GDP growth. The analysis shows that this is possible, if i) only the survey questions which are best correlated with the sectoral reference series and GDP are used (pre-selection) and ii) information on the pervasiveness of changes in the survey data is used to amplify the cyclical signal. The latter approach follows the logic that changes in the average score of the survey questions should be taken more "seriously" if they are broad-based, i.e. reflected by many questions. This amplification of pervasive short-term changes in the component survey series facilitates to adequately mimic developments in quarter-on-quarter GDP growth.

Based on that finding, the note continues to examine the nowcasting merits of the experimental indicator. A model using the new indicator as only predictor variable is shown to produce nowcasts of much higher precision than a simple auto-regressive model. When relevant hard data is added, the performance can be even further enhanced. While growth of the Industrial Production Index allows a big leap forward in nowcasting power, a lagged measure of the volatility on the European stock markets (VStoxx (-1)) has proved to improve the model even further. Generally, the added value of the experimental sentiment indicator for the purpose of nowcasting is most pronounced in the crisis period (2008q2 – 2009q4). The hard data seems to exert the biggest added value in the post-crisis period (2010q1 to 2013q2).