The predictive power of the European Economic Sentiment Indicator

Sarah Gelper and Christophe Croux

The Predictive Power of the European Economic Sentiment Indicator

Sarah Gelper and Christophe Croux

January 22, 2007

Abstract: Economic sentiment surveys are carried out by all European Union member states on a monthly basis. The survey outcomes are used to obtain early insight into future economic evolutions and often receive extensive press coverage. Based on these surveys, the European Commission constructs an aggregate European Economic Sentiment Indicator (ESI). This paper compares the ESI with more sophisticated aggregation schemes based on two statistical methods: dynamic factor analysis and partial least squares. We compare the aggregate sentiment indicators and the weights used in their construction. Afterwards a comparison of their forecast performance for two real economic series, industrial production growth and unemployment, follows. Our findings are twofold. First it is found that the ESI, although constructed in a rather ad hoc way, can compete with the indicators constructed according to statistical principles. Secondly, the predictive power of the sentiment indicators, as tested for in an out-of-sample Granger causality framework, is limited.

1 Introduction

Every month, the European Commission publishes the European *Economic Sentiment Indicator* (ESI). This indicator is based on surveys that aim to get insight into the beliefs of the economic agents. The basic idea behind the ESI is that, if consumers and manufacturers feel confident about the current and future general economic and own financial situation, they are more willing to increase their consumption and production respectively. In this sense, the ESI can be considered as an early indication of future economic activity in Europe, and hence be used as a guideline for both businessmen and policy makers. In the spirit of an integrated European market, we study whether a European sentiment indicator has informative content concerning the present and future state of economic activity in Europe.

The ESI is based on sentiment surveys carried out in all member states of the European Union. There are four business surveys, one for each of the industrial,

service, construction and retail sector, and one consumer survey. For every country, a subset of 15 questions out of these surveys is used in the construction of the ESI. This results in a large number of series which are summarized into an aggregate sentiment indicator by taking a linear combination of them. Given the increasing integration of the European economy, such an aggregate indicator might be more informative for future economic activity in each country than the national indicators are. In this paper, we compare the construction of the ESI with two other aggregation schemes and study their usefulness for predicting industrial production and unemployment at different forecast horizons. The latter are two important economic quantities, for which data availability is rather slow as compared to the survey results. As such, if the aggregate sentiment indicators give early information on quantities like industrial production and unemployment, they are highly relevant for both businessmen and policy makers.

For the ESI, as constructed by the European Commission, the weights of the components depend on intuitive economic reasoning (more details are given in Section 2). The research question in this paper is whether the construction of the ESI can be improved. In other words, we investigate whether other aggregation methods, using the same sentiment components, may result in more informative indicators. In particular we compare the ESI with sentiment indicators obtained by data driven aggregation methods, namely the *dynamic factor model* as in Stock and Watson (2002), and the *partial least squares* approach. We compare these two indicators with the ESI in three respects: (i) the evolution of the indicators over time, (ii) the importance given to European countries and the 5 surveys in the aggregation scheme, and (iii) their predictive power for industrial production growth and unemployment rates.

Given the continuously increasing availability of large amounts of economic time series, the problem of summarizing a large number of time series in one indicator is gaining popularity in the recent economic literature. For example, Cherchye et al. (2006) propose a composite indicator to capture the performance of the individual EU members in the evolution towards a single integrated European market in one single number. In a forecasting context, Banerjee et al. (2005) and Marcillino et al. (2003) find that the use of a dynamic factor model constructed from many economic indicators (but not including sentiment indicators) improves forecasts of aggregate European real economic variables as compared to univariate modeling. Hansson et al. (2005) study the forecasting performance of business survey data in Sweden, also using a dynamic factor model and find good results for forecasting GDP growth. A related study focussing on sentiment indicators is Slacalek (2005). He applies a dynamic factor model to individual questions of the Michigan sentiment survey. The resulting factors are found to be a stable predictor of US consumption growth. Our study distinguishes from Slacalek (2005) since we explicitly compare different aggregation schemes and their out-of-sample forecasting performance. Furthermore, we do not limit our attention to consumer sentiment but combine it with results from production surveys, and work in a European context.

The sentiment surveys carried out by the European Commission are not only

used in the construction of a European aggregate ESI, but also for national sentiment indicators. The predictive power of these national indicators is addressed in numerous studies (e.g. Lemmens et al. (2005), and the references therein), and results in mixed findings strongly depending on whether an in-sample or out-of-sample testing framework is used. A recent article by Cotsomitis and Kwan (2006) finds that the out-of-sample evidence for the forecasting power of national ESI and consumer confidence indicators for household spending is very limited. The research question addressed by this paper is to find out whether it is possible to improve the forecasting performance of the European ESI, but using different aggregation schemes to construct it. It will turn out that the ESI, although constructed in a rather ad hoc way and based on economic intuition, can compete with the indicators constructed according to statistical principles.

The remainder of this paper is organized as follows. Section 2 first clarifies how the European Commission constructs the ESI and then briefly explains the indicators based on the factor model and on the partial least squared method. A detailed comparison of these three indicators can be found in Section 3. Section 4 outlines the framework to test for the predictive power and compares the forecast performance of the aggregate sentiment indicators. Finally, Section 5 concludes.

2 Construction of a European Aggregate Sentiment Indicator

The aim of constructing an aggregate indicator is to catch most information contained in a large number of series into one single indicator series. In our setting, each original series corresponds to a certain question from one of the 5 sentiment surveys, and this for every member state of the EU. Table 1 presents the 15 sentiment components, resulting from the 5 different surveys. Each component corresponds to a survey question, and is expressed in *balance*, i.e. the percentage of positive answers minus the percentage of negative answers on that question for a particular country. We work with surveys from 15 European countries, the member states of the EU in 1995, resulting in a total of 15 times 15, i.e. 225, time series. Our aim is to find a method to summarize these 225 series in one indicator which can be interpreted as reflecting the general European economic sentiment. There already exist several methods to this end. Here, three of them are considered: the methodology used by the European Commission to construct the ESI, the dynamic factor model (DF) and the partial least squares (PLS) method. These three methods have in common that they construct a linear combination of the original sentiment series, but differ in the way the weights for the linear combination are calculated.

The first aggregation method is used by the European Commission to construct the ESI and proceeds in two steps. In a first step, each component is aggregated over the member states using specific country-weights. Then, in a second step these 15 component series are aggregated by making use of survey weights to end up with

one single indicator. For a component j, the weight of country i for month t is denoted by $w_{i,j,t}$, and given as a two year moving average by

$$w_{i,j,t} = \frac{v_{i,j,t} + v_{i,j,t-12}}{2} \quad \text{with} \quad v_{i,j,t} = \frac{X_{i,j,t}}{X_{EU,j,t}}. \tag{1}$$

Here $X_{i,j,t}$ is a certain economic variable measured for member state i at time t and $X_{EU,j,t}$ is the European equivalent. These variables are assumed to remain constant over one year. The economic variable $X_{i,j,t}$ differs according to the survey from which component j originates. For the industrial, construction and services sentiment surveys, $X_{i,j,t}$ is the gross value added at constant prices in the respective sector for country i at time t. For retail and consumer sentiment, $X_{i,j,t}$ represents the private final consumption expenditure at constant prices for country i at time t. From equation (1) it is clear that the country weights are time varying. In practice, they only change very slowly over the years. A weighted sum over the index i yields the value the sentiment components at the EU level at time t.

After having obtained the 15 EU-level sentiment components, these are aggregated using survey weights. These weights are given in the last column of Table 1 and based on two criteria[1]. First, they should reflect the importance of the corresponding sector in the total economy. For instance the service sector is responsible for a larger amount of total GDP than the retail sector, and therefore gets a larger weight. Second, the more the survey results from a certain sector co-move with GDP, the more weight this survey should get. Taking these two criteria into account, the European Commission decided on the following weights for the 5 surveys: 40 % for industrial confidence, 30 % for services, 5 % for retail trade, 5% for construction and 20% for consumer confidence. The weight of each survey is then equally divided over the different questions j within the survey, as illustrated in 1.

As a second way for constructing an aggregate indicator, we use the methodology of the dynamic factor model as described in Stock and Watson (2002). One factor is extracted from the 225 sentiment component series by making use of the method of principle components. This method allows to extract several underlying factors from a large number of series. Here we restrict ourselves to only one factor that is compared to the ESI. This is in accordance with the idea that there is one driving force behind economic sentiment in all countries. The first factor derived from a large data set is that linear combination of the individual series that maximizes the variance of the factor, subject to the constraint that the sum of all squared weights equals one. For a profound discussion, see Stock and Watson (2002). As opposed to the weights for the ESI, the determination of the weights in the factor analysis is solely based on the past values of observed sentiment component series.

As a third methodology for reducing the dimension of the 225 series into one single series, we consider partial least squares, a popular technique in engineering and chemistry for prediction problems with many explicative variables. We consider

---

[1]User guides with more detailed information are provided by the European Commission on the web page http://ec.europa.eu/economy_finance/indicators/businessandconsumersurveys_en.htm

	Sentiment component	Component weight
	Business Survey in the Industrial Sector	40%
1.	Assessment of order-book levels	13.3%
2.	Assessment of stocks of finished products	13.3%
3.	Production expectations for the months ahead	13.3%
	Business Survey in the Services Sector	30%
4.	Business situation of recent months	10%
5.	Evolution of demand in recent months	10%
6.	Evolution of demand expected in the months ahead	10%
	Business Survey in the Retail trade Sector	5%
7.	Business activity over recent months	1.7%
8.	Assessment of stocks	1.7%
9.	Expected business activity	1.7%
	Business Survey in the Construction Sector	5%
10.	Assessment of order books	2.5%
11.	Employment expectations for the months ahead	2.5%
	Consumers' Survey	20%
12.	Financial situation over next 12 months	5%
13.	General economic situation over next 12 months	5%
14.	Unemployment expectations over next 12 months	5%
15.	Savings over next 12 months	5%

Table 1: The fifteen sentiment components with their weights used in the construction of the European Economic Sentiment indicator. Each component corresponds to a question asked in one of the 5 different surveys.

this method in the spirit of investigating the predictive power of European sentiment indicators. While neither the factor model nor the ESI construction methods take the variable to predict into account, the method of partial least squares does. Hence PLS will construct another indicator for every economic variable to predict. The weights of the sentiment component series are chosen such that the *Covariance* between the aggregate indicator and the variable to predict is maximized. As such, the resulting indicator takes the covariance with the variable to predict into account, and is not solely based on the predictor variables, hereby aiming at a better forecasting performance. A review on PLS, with additional references, can be found in Wold (2006) and Helland (2006).

3   Comparison of the Sentiment Indicators

In a first stage of our empirical analysis, we compare the evolution of the three proposed aggregate time series: ESI_t as constructed by the European Commission, DF_t obtained from dynamic factor analysis and PLS_t resulting from the partial least squares method. We use survey question data dealing with all 15 states which were member of the European Union before the enlargement in 2004, in principle resulting in a total of 225 sentiment series (as discussed in Section 2). However, in a time range of April 1995 to November 2005, some series are incomplete and we end up with a data set of 160 series. These 160 series are used to construct common indicators based on factor analysis and partial least squares. Recall that PLS_t depends on the series to predict, and in this Section it is chosen with the aim at predicting industrial production growth at the European level.

Figure 1 shows the evolution of the indicators from April 1995 until November 2005. The dynamic factor and partial least squares method are applied to the component series in differences, as their levels are borderline non-stationary. The stationarity condition is required for consistency of the dynamic factor model, see Stock and Watson (2002). Furthermore, the differenced sentiment series are as informative as the series in levels, and the most important information is to know by which amount the general sentiment increases or decreases. Though, the sentiment in levels is more appealing for graphical representation and can easily be reconstructed from the differences, as in Figure 1. The DF and PLS sentiment series are recursively obtained: at each time point t, the indicator is extracted from all sentiment component series up to moment t, using a starting up period of 40 months. Only information from the past is included in the calculation of the current weighting scheme, so that the indicators are computed at real time. It follows from this updating procedure that the DF- and PLS-weights of the sentiment component series are time varying.

Figure 1 shows that the three indicators closely move together, even though the ESI is not based on formal statistical arguments like the other two indicators. These latter have an instantaneous correlation with the ESI of about 95%, while also important cross-correlations at different leads and lags are present. One should
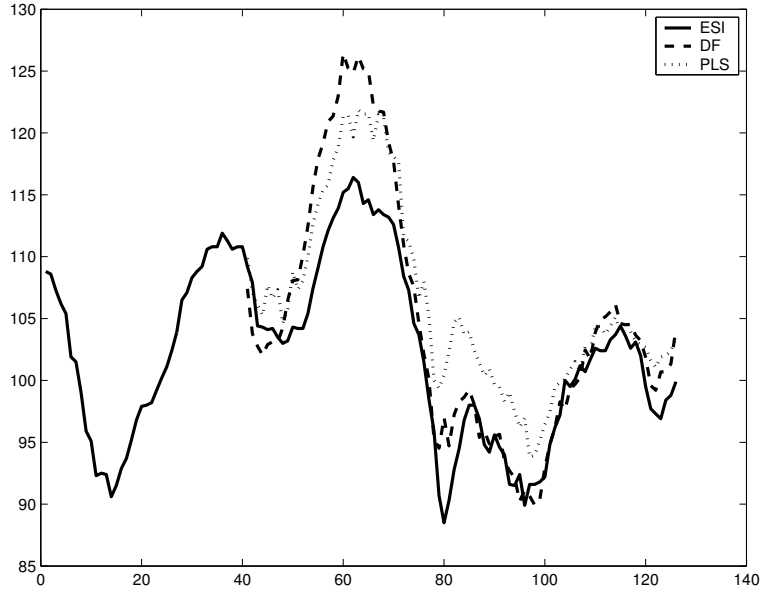
Figure 1: Three sentiment indicators: ESI_t, DF_t and PLS_t, from November 1995 until January 2006.

however be aware of the fact that all series in Figure 1 are close to being unit-root processes[2], which may result in spurious correlations. In fact, the Johansen test reveals that the three series are cointegrated (p-value of 0.01). The correlation of between the differenced series equals 48% for ESI_t and DF_t, and 38% for ESI_t and PLS_t.

Apart from comparing the sentiment indicators as such, it is also interesting to compare the weights of the component series assigned by the three methods. By adding up the weights of the 15 questions for each country, we obtain country weights. Similarly, we summed the weights over the countries for each survey, resulting in the survey weights. Figure 2 plots the country-weights of the dynamic factor approach (panel a) and the weights of the partial least squares approach (panel b), versus the ESI weights. An ordinary least squares regression line is added. In both panels, the slope of the regression line is significantly positive. The statistically founded choice of country weights, resulting from factor extraction by DF and PLS, is in line with the more intuitive and economic arguments for the ESI weights. For example, Germany is a large country and has always been considered as an important member of the European Union. Accordingly, it has the highest weight in the construction of all three indicators. The same reasoning holds for France. On the other hand, Belgium for instance is much smaller and gets a low weight in the construction of the ESI. It is, however, an important country according to the DF and PLS methods. The reason is that Belgium has a very open economy, with a lot of export to neighboring countries. The outcomes for Belgian surveys are more

---

[2]Augmented Dickey-Fuller tests yield p-values of 0.42 for the ESI, 0.17 for DF and 0.49 for PLS.
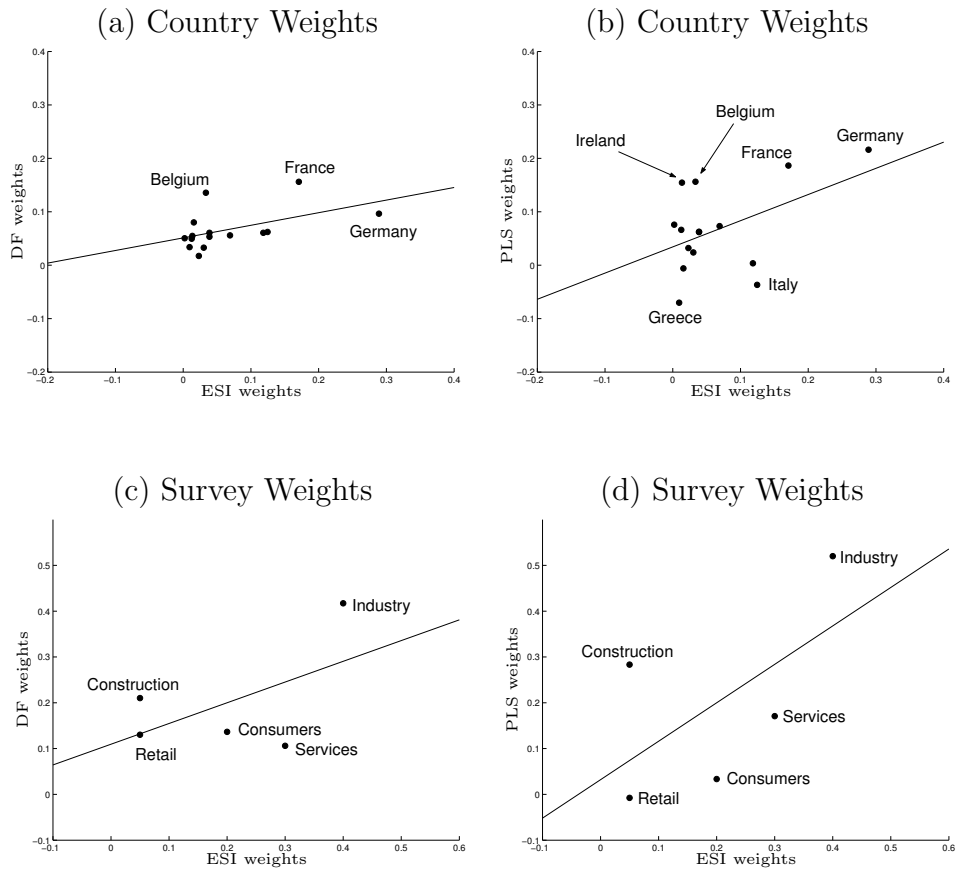
Figure 2: Scatter plots of the country weights as obtained by dynamic factor analysis (DF, panel a) and Partial Least Squares (PLS, panel b) versus ESI. Scatter plots of the survey weights as obtained by dynamic factor analysis (DF, panel c) and Partial Least Squares (PLS, panel d) versus ESI. An ordinary least squares regression line is fitted.

informative than one would expect from its country size. The PLS method also attaches much more weight to Ireland as the ESI. On the other hand, PLS gives much less weight to Greece and Italy as compared to the ESI, in fact these weights are even negative. This suggests an atypical behavior of these countries when it comes down to predicting European level industrial production growth. Panels (c) and (d) in Figure 2 compare the survey weights. We see a positive correlation between the weights used for the ESI and both the weights from DF and PLS. Especially the industrial sector gets high weighting, it represents a large percentage of the total European GDP.

8

4 Forecasting Using Sentiment Indicators

4.1 Methodology

We compare the forecasting performance of the ESI with the indicators obtained by dynamic factor analysis and the PLS method within a Granger causality framework. This allows to measure the incremental predictive power of the indicators with respect to the own past of the series to predict. Let y_t denote the series to predict and S_t the European sentiment indicator at time $t = 1 \ldots T$, where S_t is either the ESI in differences or the indicator derived from factor analysis or partial least squares based on the differenced sentiment components.

Suppose we want to forecast y_{T+h}, where h stands for the horizon of the forecast. If we take $h = 0$, we make a *nowcast* of the current y_t. Making nowcasts for macroeconomic variables is extensively discussed in Domenico et al. (2006). Many real macro-economic variables are only available with a certain time lag k, mostly of about two months. The sentiment survey results, on the contrary, are available much earlier. At the beginning of the month, we posses the survey results concerning the previous month. Therefore, the variable S_t is included up to one period ago in the forecasting model:

$$y_{t+h} = \alpha + \beta_k y_{t-k} + \gamma_1 S_{t-1} + \ldots + \beta_p y_{t-p} + \gamma_p S_{t-p} + \varepsilon_t, \tag{2}$$

with $1 \leq t \leq T$, and where k stands for the timeliness of the data-availability of the series to predict. The forecasting performance of the above model (2) is then compared with the reference model

$$y_{t+h} = \alpha + \beta_k y_{t-k} + \ldots + \beta_p y_{t-p} + \varepsilon_t, \tag{3}$$

where the sentiment indicator S_t is not used. By comparing (2) and (3), it is possible to measure how much incremental power the sentiment indicator has, corresponding to the concept of Granger causality. In the empirical application, the order p of the models is selected by minimizing the Schwarz-Bayes information criterion.

To assess the predictive content of the summary indicators, one can use in-sample as well as out-of-sample model comparison procedures. The in-sample procedure tests whether the coefficients $\gamma_1, \ldots, \gamma_p$ are jointly significant or not, while the out-of-sample procedure tests whether forecast errors using model (2) are significantly smaller than these using (3). In-sample tests are very standard and perform an F-test for the hypothesis of nullity of the γ- coefficients. The model is estimated using the entire data set, implying that in-sample tests could be subject to the overfitting problem, and may yield overoptimistic results, i.e. detect significant forecasting power while in fact there is none. Here, we only report results for the out-of-sample tests.

Out-of-sample procedures on the other hand make forecasts of observations using information from the past only and evaluate different models by comparing their "real time" forecast errors. These are measured by a recursive forecasting scheme: The first R observations are used to forecast observation $R + h$, after which the

forecast is compared with the realized value. This yields a first forecast error. Then, observation $R+1+h$ is forecasted using all observations up to period $R+1$ and again the forecast and the realized value are compared. This procedure continues until the end of the series and results in a series of h-step-ahead forecast errors. We selected $R = T/2$, resulting in a sequence of $T/2 - h$ forecast errors, which are summarized by a Mean Squared Forecast Error (MSFE). The MSFE of model (2) is compared with that of model (3) and if the decrease in MSFE is significant, then we say there is a significant (incremental) forecasting power of the aggregate sentiment indicator S_t. For comparing the MSFEs we use the encompassing regression test statistic, which is described and found to be well-performing in Clark and McCracken (2001). The associated p-values are obtained by a residual bootstrap procedure.

To remain in a pure out-of-sample framework, it is necessary to recompute the indicators obtained from the dynamic factor analysis or partial least squares method in each step. More precisely, when forecasting of observation $R + s + h$, in step s of the recursive scheme, the weights for the 225 sentiment series are computed based on the first $R + s$ observations only. Out-of-sample tests require much computation time, but are conceptually more natural since they mimic the process of true real time predictions of future values.

In the next subsections, our aim is to find out whether the ESI helps in forecasting real industrial production or unemployment rates, and secondly to see how well the ESI performs relative to the methods of dynamic factor analysis and partial least squares. We work with industrial production and unemployment data at the aggregate European level, and at the country-level for Belgium, Denmark, Germany, France, Italy, Luxembourg and the UK. The selection of these eight countries is mainly based on data availability, yielding a complete data set which ranges from April 1995 to November 2005. These data are collected by the European Commission and can be downloaded from the Eurostat website[3].

4.2 Forecasting Industrial Production

In this section we test for the forecasting performance of the different European indicators for predicting industrial production growth. We use using the out-of-sample testing procedure for Granger Causality outlined in Section 4.1. The industrial production index in levels contains a stochastic trend, but is stationary in differences for all countries and for the European aggregate (according to the augmented Dickey-Fuller test). Therefore, we estimate models (2) and (3) with y_t the real industrial production index in differences. For the same reason, S_t will stand for the common sentiment indicator in differences.

We make nowcasts and forecasts for one, three and six months ahead, and the timeliness k equals 2 for the industrial productions series. Table 2 presents the p-values for the out-of-sample Granger causality tests, where the null hypothesis states that there is no incremental predictive power. Results are given for the three

---

[3]ec.europa.eu/eurostat

	h = 0			h = 1			h = 3			h = 6		
	ESI	DF	PLS	ESI	DF	PLS	ESI	DF	PLS	ESI	DF	PLS
Belgium	0.85	0.78	0.90	0.79	0.84	0.50	0.29	0.48	0.75	0.81	0.23	0.74
Denmark	0.41	0.13	0.65	0.15	0.91	0.93	0.09	0.40	0.06	0.90	0.48	0.65
Germany	0.29	0.79	0.06	0.06	0.23	0.22	0.04	0.30	0.00	0.18	0.42	0.00
France	0.43	0.55	0.08	0.21	0.41	0.79	0.76	0.37	0.11	0.95	0.34	0.52
Italy	0.50	0.88	0.00	0.89	0.52	0.45	0.90	0.11	0.02	0.03	0.01	0.00
Luxembourg	0.45	0.10	0.65	0.25	0.45	0.49	0.90	0.72	0.68	0.22	0.50	0.67
Netherlands	0.75	0.28	0.14	0.32	0.54	0.96	0.78	0.31	0.16	0.43	0.43	0.41
UK	0.41	0.48	0.30	0.37	0.08	0.00	0.01	0.27	0.08	0.10	0.09	0.00
Europe	0.55	0.68	0.06	0.01	0.03	0.15	0.63	0.83	0.86	0.96	0.70	0.19

Table 2: *Forecasting Industrial Production Growth: P*-values for Out-Of-Sample Granger-causality tests at differen horizons h for eight countries and the European aggregate, using the ESI or the indicators obtained by the Dynamic Factor model (DF) of the Partial Least Squares method (PLS).

common sentiment indicators, the ESI and the ones resulting from the DF and PLS methods, for industrial production growth in eight European countries as well as for the European aggregate. From Table 2 it becomes clear that the ESI is of limited use for making forecasts of industrial production. In some specific cases, a significant causality relation of the common indicators for the industrial production is found, more precisely for the aggregate European production at the one-month-ahead horizon and for the 6 months ahead Italian industrial production. However, one should keep in mind that Table 2 represents a total of 36 statistical test for the ESI, hence one could expect about two cases where we find significant effects, even when no Granger Causality is present. We conclude that overall the ESI has limited out-of-sample forecasting power. This conclusion is in line with the findings of Cotsomitis and Kwan (2006).

Let us now compare the performance of the ESI relative to the other two indicators, based on the DF and PLS methods. The PLS method seems to perform best, yielding the highest number of significant outcomes, in particular for the country specific variables. Overall, however, we may conclude that also the DF and PLS approaches result in indicators with limited forecasting ability, and that they do not yield a clear improvement in forecasting power with respect to the ESI. When we focus attention on the prediction of the aggregate European industrial production, there is one outcome that is consistently found to close to significance. At a horizon of one month, the sentiment indicators detect a Granger causal relationship. This may be due to the fact that question 1 of Table 2 is directly related tot the variable of interest: the industrial production within one month will directly depend on the order book level.

	h = 0			h = 1			h = 3			h = 6		
	ESI	DF	PLS	ESI	DF	PLS	ESI	DF	PLS	ESI	DF	PLS
Belgium	0.34	0.16	0.81	0.32	0.10	0.96	0.95	0.75	0.85	0.00	0.32	0.40
Denmark	0.20	0.44	0.80	0.26	0.50	0.42	0.55	0.60	0.67	0.70	0.60	0.82
Germany	0.24	0.27	0.20	0.08	0.02	0.02	0.74	0.05	0.00	0.81	0.84	0.00
France	0.26	0.54	0.98	0.69	0.78	0.26	0.45	0.29	0.27	0.01	0.00	0.46
Italy	0.09	0.40	0.60	0.09	0.14	0.94	0.10	0.01	0.47	0.26	0.50	0.70
Luxembourg	0.76	0.65	0.60	0.35	0.19	0.62	0.15	0.18	0.73	0.98	0.97	0.91
Netherlands	0.02	0.73	0.11	0.27	0.38	0.07	0.51	0.85	0.30	0.19	0.07	0.40
UK	0.64	0.17	0.63	0.80	0.86	0.34	0.94	0.51	0.27	0.15	0.29	0.89
Europe	0.03	0.89	0.97	0.24	0.54	0.39	0.63	0.31	0.46	0.01	0.04	0.94

Table 3: *Forecasting Unemployment Rates: P*-values for Out-Of-Sample Granger-causality tests at differen horizons h for eight countries and the European aggregate, using the ESI or the indicators obtained by the Dynamic Factor model (DF) of the Partial Least Squares method (PLS).

4.3   Forecasting Unemployment Rates

To extend the conclusions made in Section 4.2, we conduct the same forecasting exercise, but for predicting unemployment rates. Unemployment is an important economic indicator and is strongly linked with positive and negative feelings towards the general economy. Entrepreneurs who fear a decline in sales will downsize the personnel, and a high unemployment rate makes consumers more pessimistic. Therefore, we expect the sentiment surveys to be informative about unemployment rates. A preliminary in-sample study confirms this expectation. Though, we are more interested in out-of-sample performance and these results are reported in Table 3. As before, we see that in general there is no or only weak evidence of Granger Causality. When we compare the ESI with the other two common indicators, there is no clear preference for one of the three weighting schemes, none of the methods clearly outperforms the others (here, the DF method gives the largest number of significant outcomes). In some cases there is evidence of predictive power, for example for German unemployment at $h=1$, of for France at a half year horizon. For the European aggregate, both the ESI and DF find significant predictive power (but PLS not at all), at the larger forecast horizon of six months.

5   Conclusion

This paper compares the European Economic Sentiment Indicator, as published monthly by the European Commission, with two other methods for constructing an aggregate sentiment indicator. The two alternative ways of aggregating the 225 sentiment component series are based on pure statistical techniques: a dynamic factor analysis and the partial least squares method. In Section 3 we compare

these three different indicators. The weights used in aggregating the many single component series (being question and country-specific) are shown to be related, though far from identical. The evolution of the 3 different indicators is, on the other hand, very similar (Figure 1). Section 4 compares the relative performance of the indicators for predicting two real economic variables. Their performance is again similar. We conclude that although the European Economic Sentiment Indicator seems to be constructed in a rather ad hoc way, it is quite comparable with other construction schemes based on statistical arguments.

Furthermore, we find that the aggregate sentiment indicators are weak predictors of industrial production and unemployment. Granger Causality is only found at some particular horizons and the ESI, dynamic factor model and partial least squares method behave comparably. This indicates that the weak forecasting power of the ESI is not due to the way it is constructed, but simply comes from the fact that the individual component series, even when we combine them, do not contain enough information for making accurate predictions. In other words, anticipations of consumers and producers about the future state of the economy may not be informative enough for predictive purposes.

Recall that all tests for forecasting accuracy were done in an out-of-sample framework. We also did the corresponding in-sample tests, which indicate a slightly stronger, but still limited, presence of Granger Causality. It is interesting to note that when previous studies reported significant Granger Causality for national or other sentiment indicators, they always used an in-sample testing framework (e.g. Carroll et al. (1994), Desroches and Gosselin (2002), Bryant and Macri (2005)). Our believe, however, is that an out-of-sample framework is a better proxy for a true forecast exercise. Another possibility, taken by several authors, is not to measure the incremental forecasting power of the sentiment indicator, but only look for direct causality, by excluding the lagged values of the variable to predict. We also carried out the out-of-sample tests for direct causality, but they only gave a marginal improvement in significance with respect to the Granger causality results.

References

Banerjee, A., Marcillino, M., Masten, I., 2005. Leading indicators for euro-area inflation and GDP growth. Oxford Bulletin of Economics and Statistics 67, 785–813.

Bryant, W., Macri, J., 2005. Does sentiment explain consumption? Journal of economics and finance 29, 97–111.

Carroll, C., Fuhrer, J., Wilcox, D., 1994. Does consumer sentiment forecast household spending? if so, why? The American Economic Review 84, 1397–1408.

Cherchye, L., Knox Lovell, C., Moesen, W., Van Puyenbroeck, T., 2006. One market, one number? A composite indicator assessment of EU internal market dynamics. European Economic Review, Article in press.

Clark, T., McCracken, M., 2001. Tests of equal forecast accuracy and encompassing for nested models. Journal of Econometrics 105, 85–110.

Cotsomitis, J., Kwan, C., 2006. Can consumer confidence forecast household spending? Evidence from the European Commission busniness and consumer surveys. Southern Economic Journal 72, 597–610.

Desroches, B., Gosselin, M., 2002. The usefulness of consumer confidence indexes in the united states. Bank of Canada working paper.

Domenico, G., Reichlin, L., Small, D., 2006. Nowcasting GDP and inflation: the realtime informational content of macroeconomic data releases. Working Paper Series from European Central Bank 633.

Hansson, J., Jansson, P., Lof, M., 2005. Business survey data: do they help in forecasting GDP growth? International Journal of Forecasting 21, 377–389.

Helland, I., 2006. Partial least squares regression. In: Kotz, S., Read, B., Balakrishnan, N., Vidakovic, B. (Eds.), Encyclopedia of Statistical Sciences, 16 Volume Set, 2nd Edition. pp. 5957–5962.

Lemmens, A., Croux, C., Dekimpe, M., 2005. On the predictive content of production surveys: a pan-european study. International Journal of Forecasting 21, 363–375.

Marcillino, M., Strock, J., Watson, M., 2003. Macroeconomic forecasting in the euro area: Country specific versus area-wide information. European Economic Review 47, 1–18.

Slacalek, J., 2005. Analysis of indexes of consumer sentiment, working paper. German Insititute for Economic Research.

Stock, J., Watson, M., 2002. Forecasting using principle components from a large number of predictors. Journal of the American Statistical Association 97, 1167–1179.

Wold, H., 2006. Partial least squares. In: Kotz, S., Read, B., Balakrishnan, N., Vidakovic, B. (Eds.), Encyclopedia of Statistical Sciences, 16 Volume Set, 2nd Edition. pp. 5948–5957.