Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics

Document de Treball 2021/03, 27 pàg.
Working Paper 2021/03, 27 pag.

Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group

Document de Treball 2021/01, 27 pàg.
Working Paper 2021/01, 27 pag.

"Nowcasting and forecasting GDP growth with machine-learning sentiment indicators"

Oscar Claveria, Enric Monte and Salvador Torra

UBIREA
Research Institute of Applied Economics

AQR
Regional Quantitative Analysis Research Group

Abstract

We apply the two-step machine-learning method proposed by Claveria et al. (2021) to generate country-specific sentiment indicators that provide estimates of year-on-year GDP growth rates. In the first step, by means of genetic programming, business and consumer expectations are evolved to derive sentiment indicators for 19 European economies. In the second step, the sentiment indicators are iteratively re-computed and combined each period to forecast yearly growth rates. To assess the performance of the proposed approach, we have designed two out-of-sample experiments: a nowcasting exercise in which we recursively generate estimates of GDP at the end of each quarter using the latest survey data available, and an iterative forecasting exercise for different forecast horizons We found that forecasts generated with the sentiment indicators outperform those obtained with time series models. These results show the potential of the methodology as a predictive tool.

JEL Classification: C51, C55, C63, C83, C93.

Keywords: Forecasting, Economic growth, Business and consumer expectations, Symbolic regression, Evolutionary algorithms, Genetic programming.

Oscar Claveria: AQR–IREA, Department of Econometrics, Statistics and Applied Economics, University of Barcelona, Diagonal 690, 08034 Barcelona, Spain. Tel.: +34-934021825. Email: oclaveria@ub.edu

Enric Monte: Department of Signal Theory and Communications, Polytechnic University of Catalunya (UPC). Email: enric.monte@upc.edu

Salvador Torra: Riskcenter–IREA, Department of Econometrics, Statistics and Applied Economics, University of Barcelona (UB). Email: storra@ub.edu

1. Introduction

The pandemic has caused a disruption in the evolution of macroeconomic aggregates. Consequently, the estimation of upcoming events becomes one of the fundamental objectives of economic analysis, especially in periods of high uncertainty such as the current one. Recent trade disputes and growing investor concerns about the global economic outlook have led the International Monetary Fund (IMF) to downgrade global growth projections for 2020, which have their lowest levels since the 2008 financial crisis (IMF, 2020). In this context, agents' expectations about future economic conditions are a key feature in macroeconomic forecasting.

Expectations are not directly observable. Consequently, agents' expectations tend to be elicited via surveys. Survey expectations present several advantages over experimental expectations: (a) they are based on the knowledge of respondents operating in the market, (b) they provide detailed information about a wide range of economic variables, and (c) they are available ahead of the publication of official quantitative data. These features make them very useful for prediction.

One of the main sources of expectation information are economic tendency surveys (ETS). In ETS respondents are asked whether they expect variables to rise, fall, or remain unchanged. Some of the most well-known ETS are collected by the University of Michigan, the Federal Reserve Bank of Philadelphia, the Organisation for Economic Co-operation and Development (OECD), and the European Commission (EC). In 1961, the EC launched the Joint Harmonised Programme of Business and Consumer Surveys with the aim of unifying the survey methodologies in the member states of the European Economic Community – now the European Union (EU), allowing comparability between countries.

Survey responses from ETS are commonly used to design composite confidence and sentiment indicators such as the ifo World Economic Climate Index, the University of Michigan Consumer Confidence Index or the Purchasing Managers' Index calculated by the Markit Group. The EC constructs business and consumer confidence indicators as the arithmetic mean of a subset of predetermined survey expectations.

The selection of variables for construction of confidence indicators is fundamentally determined by their fit to a reference series. As noted by Abberger et al. (2018), economic relationships between variables change over time and require periodic overhaul. Therefore, in this study we propose a machine-learning method for the generation of economic sentiment indicators that allows both an automated variable selection procedure and an update of the relationships between the selected variables.

The proposed approach allows determination of an optimal combination of expectations that minimises a set loss function. The obtained expressions differ from the confidence indicators constructed by the EC in that: (a) they are based on information coming from all the available variables of each survey, (b) they select expectations with the highest forecasting power and their optimal lag structure, (c) they capture the existing non-linear relationships between survey expectations, and (d) they generate direct estimates of economic growth.

The objective of the paper is threefold. First, we aim to provide practitioners with easy-to-implement business and consumer confidence indicators. To this end, we have used all the variables contained in the industrial and consumer surveys conducted by the EC for 19 EU states and for the euro area (EA). With this information, we generated country-specific confidence indicators that estimate the GDP growth rate expected by firms and consumers. Secondly, because the algorithm selects the expectational variables with the highest predictive capacity, including the number of lags, we evaluate the relative importance of the variables in each survey as well as their lag structure.

Finally, we assess the forecasting performance of the generated indicators. On the one hand, we compare them to the confidence indicators constructed by the EC in a nowcasting exercise. On the other hand, we design a recursive out-of-sample forecasting experiment in which we iteratively re-compute the indicators to predict economic growth. The obtained forecasts are then compared to univariate time series models used as a benchmark.

The proposed methodology is based on genetic programming (GP), which is a soft computing search technique based on the application of evolutionary algorithms. GP simultaneously evolves the structure and the parameters of expressions, allowing formalisation of the interactions between the variables that best fit a reference series. This approach is especially useful in situations where the exact functional form of the solution is not known in advance – such as the present one, where there is no a priori combination of survey expectations that best tracks economic growth.

GP has been successfully used as a machine learning tool for automatic problem-solving in many areas such as image processing (Harding et al., 2013), but very seldom for macroeconomic modelling and forecasting (Álvarez-Díaz, 2019, Claveria et al., 2019, 2020).

In this study, we fill this gap by applying GP to the estimation of symbolic regressions that link economic growth with survey expectations. We designed an independent experiment for each country and for each type of survey, generating confidence indicators that allowed us to independently monitor economic growth dynamics from both the demand and the supply sides of the economy.

The rest of the paper proceeds as follows: the next section describes the methodological approach and the experimental setup. In Section 3 we present the obtained sentiment indicators. In Section 4, we assess the performance of the indicators in a nowcasting exercise. In Section 5, we perform an iterative forecasting experiment. Finally, Section 6 concludes.

2. Methodology

GP is a heuristic search technique based on the evolution of programs. This optimisation approach represents programs in tree structures that learn and adapt by changing their size, shape, and composition of the models. As opposed to conventional regression analysis, which is based on a certain ex-ante model specification, GP searches for relationships between a given set of variables and evolves the functions until it reaches a solution that can be described by the algebraic expression that best fits the data.

GP simultaneously evolves the structure and the parameters of the expressions. This feature provides a quick overview of the most relevant interactions between variables and can help to identify new unknown links. As a result, due to its suitability for finding patterns in large datasets and handling complex modelling tasks, this empirical modelling approach is attracting researchers from different areas. Although GP was first used as a means to assess the non-linear interactions between price level, gross national product, money supply, and the velocity of money (Koza, 1992), applications of GP in macroeconomics have been very limited since then. See Claveria et al. (2017a) for a recent review of the application of GP to economic modelling.

Evolutionary computation is based on the application of the principles of the theory of natural selection to an iterative optimisation problem. The implementation of GP starts by the creation of an initial random population of M individuals (functions or programs), from which the algorithm selects the fittest ones (parents). In order to guarantee diversity in the population, we used size three tournament method as the strategy for the selection of parents for replacement, meaning that the best two out of three individuals randomly selected are finally mated.

Genetic operators (reproduction, crossover and mutation) are applied to the selected parents (*N*). Reproduction results in the copying of the function; crossover consists of exchanging random parts of selected pairs; and mutation involves substitution of some random part of a function with some other.

In each successive simulation (generation), a new and fitter offspring is generated. The fitness of each member of the population is evaluated by a loss function. Operations are recursively applied to the new generations until a stopping criterion is reached. The recursion stops when some individual program reaches a predefined fitness level or when the process reaches a given number of generations (Ng). The output of this process consists of the best individual function from all generations.

In our case, we generated a first random population of 70000 functions, and selected the best 10000 individuals according to the obtained mean square error. We set a maximum number of 100 generations as the termination criterion.

In this study, we implemented GP to generate composite indicators that capture optimal combinations of survey variables that best track the actual evolution of economic growth. Formally, the objective of the algorithm is to infer a functional relationship from a set of observations, such that the inferred function $f(x_i)$ is as near as possible to the reference series in the Euclidean distance sense, where index $i = 1, \dots, M$ denotes the sample size. The search process is characterised by a trade-off between accuracy and simplicity. To limit the complexity of the resulting expressions, the set of functions is restricted to the four elementary mathematical operations (addition, subtraction, product, and division). See Nicolau and Agapitos (2020) for a detailed study on the effect of the choice of function sets on the generalisation performance of symbolic regression models.

With the aim of further restricting the complexity of the resulting functional forms, we additionally introduced regularisation terms in the slope and curvature of the inferred functions. See Hastie et al. (2009) for a justification of the need to regularise. We used the Distributed Evolutionary Algorithms in Python (DEAP) developed by Fortin et al. (2012)

3. Sentiment indicators

This study matches two sources of information: official quantitative GDP data and firms' and consumers' qualitative expectations about a wide array of variables. Regarding the quantitative information, we used seasonally adjusted year-on-year growth rates of GDP provided by Eurostat. With respect to agents' expectations, we used all monthly and quarterly data from the Joint Harmonised EU Industry and Consumer surveys conducted by the EC (see Table A1 in the Appendix). Monthly survey indicators were aggregated on a quarterly basis and can be freely downloaded at the website of the EC.

The sample period goes from 2003.Q1 to 2020.Q1. The last seventeen quarters were used as the out-of-sample period to evaluate forecast accuracy. We focused on 19 European countries – Austria (AT), Belgium (BE), Bulgaria (BG), the Czech Republic (CZ), Denmark (DK), Finland (FI), France (FR), Germany (DE), Greece (EL), Hungary (HU), Italy (IT), the Netherlands (NL), Poland (PL), Portugal (PT), Romania (RO), Slovenia (SI), Spain (ES), Sweden (SE) and the United Kingdom (UK) – and the EA.

In both surveys, respondents are asked about their expectations regarding future developments and their perceptions about past and present changes. In either case, results are presented as balance series, which are obtained from the percentage of positive replies minus the percentage of negative replies. The EC publishes one composite indicator for each survey: the Industry Confidence Indicator (ICI) for the industry survey and the Consumer Confidence Indicator (CCI) for the consumer survey. Both indicators are obtained from the arithmetic mean of the balance series of a subset of questions.

In this section, we present the industry and consumer confidence indicators obtained for each country and for the EA after the evolutionary process. We ran two independent experiments for each country. In the first one, we linked GDP growth to the industry survey indicators. In the second one, we linked GDP growth to consumer survey indicators. The output of the first set of experiments are country-specific evolved industrial confidence indicators that generate estimations of firms' expectations of economic growth (Exp.IND), while the output of the second set of experiments are evolved consumer confidence indicators for each country that yield estimations of households' expectations of the evolution of economic activity (Exp.CONS). The obtained industrial and consumer confidence indicators are respectively presented in Table 1 and Table 2.

Table 1 Evolved industrial confidence indicators

Austria	$0.10 * B1_t + 0.02 * B10_{t-2} + 0.01 * B1_{t-2} + 0.01 * B2_t - 0.01 * B4_{t-4}$
Belgium	$\frac{B11_{t-1} + B11_{t-2} - B14_t * B5_{t-1} + B8_{t-3}}{B11_{t-3} + B8_t}$
Bulgaria	$10.01 * \frac{B1_t + B5_{t-2}}{B11_t}$
Czech Republic	$0.10 * \frac{-B1_{t-2} * (B8_{t-2}) + (B2_{t-2} + (B5_t - 10.01) * B8_{t-2}) * (B5_t + B5_{t-2})}{B2_{t-2} + (B5_t - 10.01) * (B8_{t-2})}$
Denmark	$0.10 * \frac{-B1_t + (B1_t + 20.01) * (B1_t + B10_{t-3})}{B1_t + 20.01}$
Finland	$0.05 * B1_t + 0.10 * B5_t + 0.05 * B9_{t-1} - 1.01$
France	$0.10 * B10_t + 0.10 * B11_{t-1} - 0.10 * B11_{t-4} + 1.01$
Germany	$\frac{0.10 * (B10_{t-1}) * (B1_t + B10_{t-1} + B14_{t-2}) + 1.01}{B10_{t-1}}$
Greece	$\frac{0.23 * B1_t * (B11_t) * (B11_{t-4}) - B11_{t-4} - (B1_{t-4} * B9_t - B1_t) * (B11_{t-4})}{(I11_t) * (I11_{t-4})}$
Hungary	$\frac{-0.51 * B1_{t-4} + 13.10 * B5_t - B6_{t-2} * B7_{t-1}}{B11_t}$
Italy	$\frac{-B1_{t-3} + B2_t + B5_{t-1} + 2.01 * B9_t}{B8_{t-3}}$
Netherlands	$0.01 * B2_{t-1} + 0.11 * B5_t + 0.10 * B5_{t-1} + 0.01 * B9_{t-1}$
Poland	$\frac{0.28 * B9_t}{0.28 * \frac{B11_{t-3}}{B3_{t-4}} + B8_{t-3} + 4.24} + 3.52$
Portugal	$0.10 * B12_t - 0.10 * B14_{t-4} + 0.10 * B4_{t-3} + 0.10 * B5_t + 0.10 * B5_{t-1} + 0.23$
Romania	$0.31 * B5_{t-1} - \frac{0.92}{\left(\frac{B2_{t-1} + 3.31}{B11_{t-4}}\right) * (B5_{t-1})}$
Slovenia	$\frac{0.10 * B11_{t-3} - 0.10 * B1_{t-1} + 0.10 * (B11_{t-3}) * (B10_{t-1} + B2_t + B4_{t-4}) - 0.05}{B11_{t-3}}$
Spain	$0.10 * B1_t + 0.10 * B4_{t-3} + 0.10 * B5_{t-2} + 0.10 * B5_{t-4} + 0.38$
Sweden	$0.10 * B5_{t-2} - 0.10 * B6_{t-1} + 0.10 * B9_t$
UK	$-0.04 * B14_{t-2} + 0.10 * B5_t + 0.41$
EA	$0.10 * B1_t + 0.10 * B10_{t-1} + 0.10 * B4_{t-2}$

Table 2 Evolved consumer confidence indicators

Austria	$\frac{2.01 * C11_t + C12_{t-2} + 2.01 * C4_{t-2} - C5_{t-3} - C9_{t-2}}{C12_{t-1}}$
Belgium	$\frac{C13_{t-3} + C14_{t-1} + C2_{t-2} * C3_{t-2}}{C10_t + C13_{t-3}}$
Bulgaria	$\frac{0.10 * C6_{t-1} * (C10_t + 2.01 * C7_t) - (C10_t + C7_t) * C14_{t-2}}{C14_{t-2}}$
Czech Republic	$0.10 * (-C15_{t-2} + C3_t + C8_t - C9_{t-4})$
Denmark	$\frac{C3_t * C7_{t-2} + (C11_{t-1}) * (C12_{t-3} + C3_t + C4_{t-3} - C5_{t-2})}{(C11_{t-1}) * (C12_{t-3})}$
Finland	$0.10 * \frac{-C15_{t-4} * (C3_{t-2}) + (C3_t + C8_{t-4}) * (-0.50 * C8_{t-4} + C10_{t-3} * C3_{t-2})}{-0.50 * C8_{t-4} + (C10_{t-3}) * (C3_{t-2})}$
France	$2.06 + \frac{0.15 * C13_{t-3}}{C3_{t-1}} - \frac{1.85 * C1_{t-4}}{C13_{t-2} - C3_{t-1}}$
Germany	$0.05 * C10_{t-2} + 0.04 * C4_{t-1} - 0.05 * C5_{t-2} - 0.05 * C7_t + 1.86$
Greece	$0.10 * C15_{t-3} - 0.10 * C2_{t-1} + 0.10 * C7_t - 1.01$
Hungary	$4.54 + \frac{0.14 * C13_t}{C9_t} + \frac{3.03 * C8_t}{C3_t - C13_{t-1}}$
Italy	$\frac{C6_{t-4} * (-C6_{t-4})}{11.01 * C6_{t-4} - C10_{t-2} + 20.50}$
Netherlands	$\frac{C14_{t-1} + C1_{t-2} + C6_{t-4} + 2.01 * C7_t + C9_{t-1}}{C13_{t-4}}$
Poland	$\frac{C14_{t-1} + C14_{t-2} + C8_{t-1} * C9_t + C7_t}{C15_{t-1}}$
Portugal	$\frac{C14_{t-1} + C14_{t-3} - C2_{t-2} + C12_{t-4} * C5_{t-3} - C12_{t-4} * C2_{t-2}}{C13_{t-4}}$
Romania	$\frac{4.65 * C10_t - C2_t + C13_{t-4} + 4.65 * C7_t}{C8_t}$
Slovenia	$\frac{C14_{t-1} - C6_{t-1} + C7_t - 2.01 * C8_t}{C11_{t-2}}$
Spain	$\frac{-5.17 * C1_t + 0.76 * C15_t - 2.01 * C3_{t-1} + 4.44 * C9_t}{C11_{t-3}}$
Sweden	$\frac{-C4_{t-1} + (C8_{t-1}) * (C4_{t-2} + C1_t + C8_{t-1})}{(C2_{t-4}) * (C8_{t-1})}$
UK	$\frac{C14_{t-3} * (C14_{t-3} - C3_t) - C14_{t-3} - C3_t * C4_{t-4} + C8_{t-1}}{(C15_t) * (C14_{t-3} - C3_t)}$
EA	$\frac{C13_{t-1} + C13_{t-3} + C14_{t-2} - C2_{t-2} * C7_t + 3.78}{C13_{t-3}}$

When comparing the resulting indicators of industrial and consumer confidence, we observed that genetic algorithms generated more linear expressions for firms' expectations. In most countries, the derived expression is a linear combination of several industry survey variables, as opposed to evolved consumer confidence indicators, which are mostly non-linear and, include ratios and more complex interactions between survey indicators.

Regarding the lag structure, most variables tend to appear indistinctly with and without lags, sometimes for the same country. In the case of the evolved consumer indicators, the financial and general economic situation over the last 12 months, as well as future unemployment expectations, mostly appear contemporaneously, in the same way as the observed production trend and new orders in recent months appear for the evolved industry indicators. The results of Table 1 and Table 2 are summarised in Fig. 1.

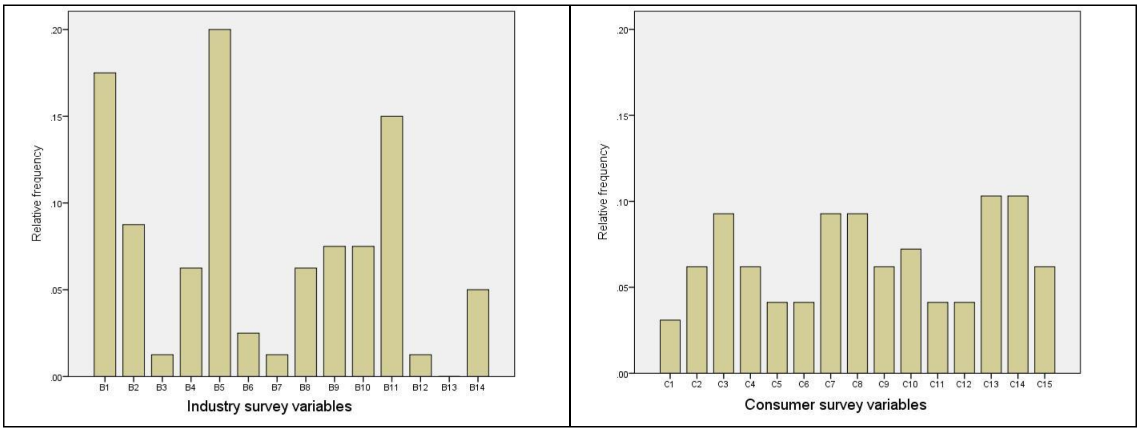

Fig. 1. Bar chart with relative frequency of variable selection (industry and consumer survey)

In the bar chart which shows the relative frequency with which each survey variable appears in the evolved expressions, we can observe that variable *B5* from the industry survey ('production expectations for the months ahead') is the most frequent of the evolved industry confidence indicators. Regarding consumer expectations, variables *C13* ('intention to buy a car') and *C14* ('intention to purchase a home in the next 12 months') are the variables most frequently selected by the algorithm, both contemporaneously and with lags. We observe that the distribution of the industry survey variables shows less variance than that of the consumer survey variables, which is more flat-topped, showing fatter tails. It can be seen that each survey variable of the consumer survey appears at least 3 or 4 times in the evolved consumer confidence indicators; however in the industry survey, production expectations appear 16 times, while other variables such as the 'competitive position inside the EU' do not appear for any country.

These obtained results suggest the predictive potential of production expectations in the industry. In the case of consumers, the intention to buy a car or a house are the variables with the highest informational content to capture economic growth dynamics. Klein and Özmucur (2010) also found evidence of the predictive potential of variable *B5* when evaluating the usefulness of expectations from the industry survey to improve the forecasting performance of time series models in 26 European countries. It is also noteworthy that in spite of the leading properties of the variables contained in the consumer quarterly surveys, which are the most frequently selected variables by the algorithm, they have always been omitted by the EC in the construction of the official consumer confidence indicators.

4. Nowcasting experiment

In this section, we examine the predictive performance of the proposed confidence indicators in tracking economic activity in two different forecasting exercises. We used the last 17 quarters (2016.Q1 to 2020.Q1) as the out-of-sample period, and the root mean square forecasting error (RMSFE) as a measure of forecast accuracy. First, we compared the forecasts obtained with the evolved confidence indicators (Exp.IND and Exp.CONS) to those obtained with the corresponding confidence indicators constructed by the EC, previously re-scaled (Cof.IND and Cof.CONS). Because the output of the evolved indicators is directly expressed as expected annual GDP growth rates, we re-scaled the indicators presented in expressions (1) and (2), by regressing the GDP growth of each country on the components of the indicators during the in-sample period (2003.Q1 to 2015.Q4).

In Fig. 2 we graphically compare the evolution of the two GP-generated indicators to that of the GDP of each country. The last seventeen quarters of the sample are used as the out-of-sample period, in which we use the results of the surveys to estimate period-to-period economic growth prior to the publication of official data.

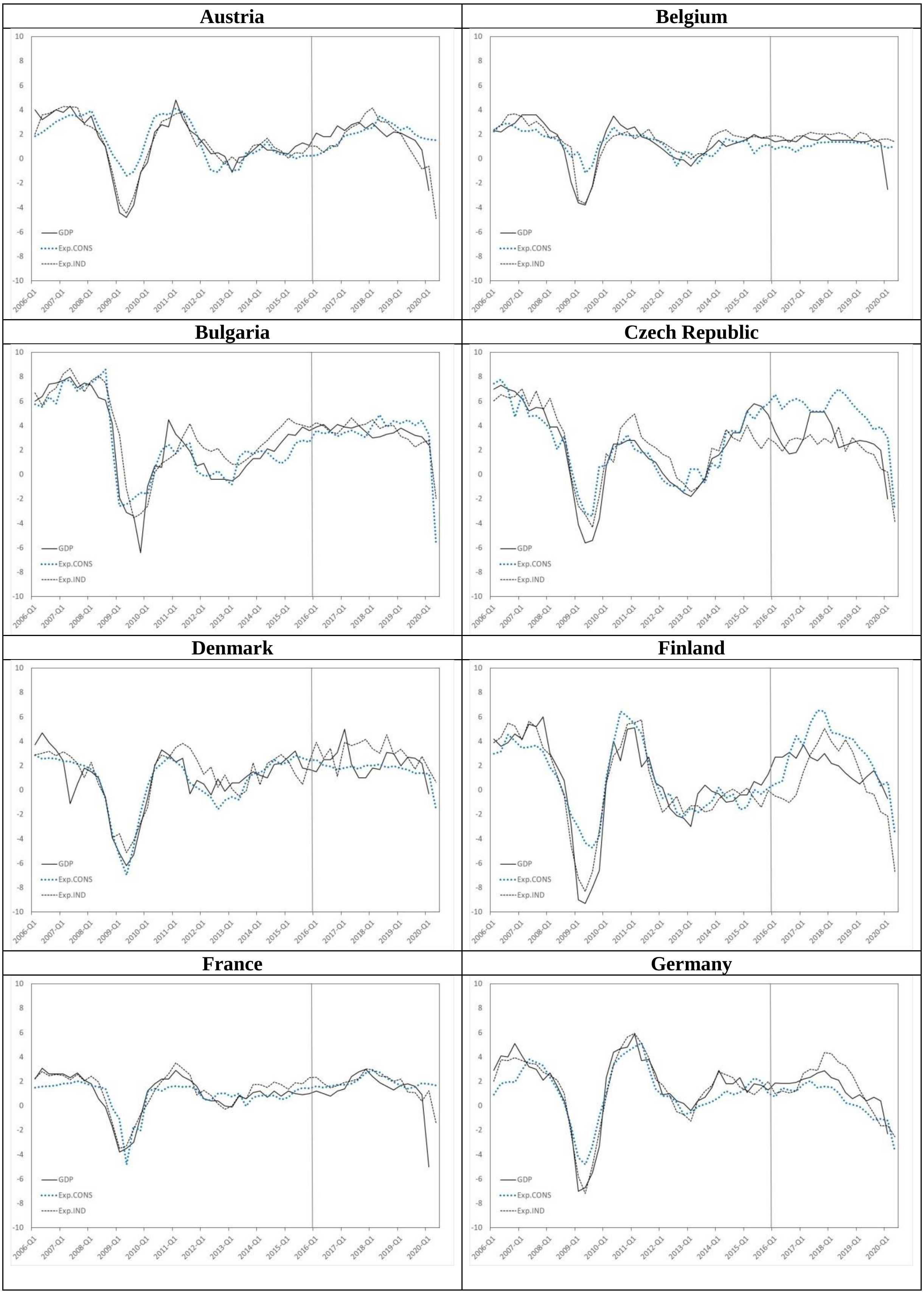

Fig. 2. Evolution of GDP and firms' and consumers' evolved confidence indicators

Notes: The black line represents the evolution of GDP growth, the grey dotted line the evolution of consumer confidence (Exp.CONS), and the dashed black line the evolution of industrial confidence (Exp.IND). The vertical line in 2016.Q1 marks the beginning of the out-of-sample period.

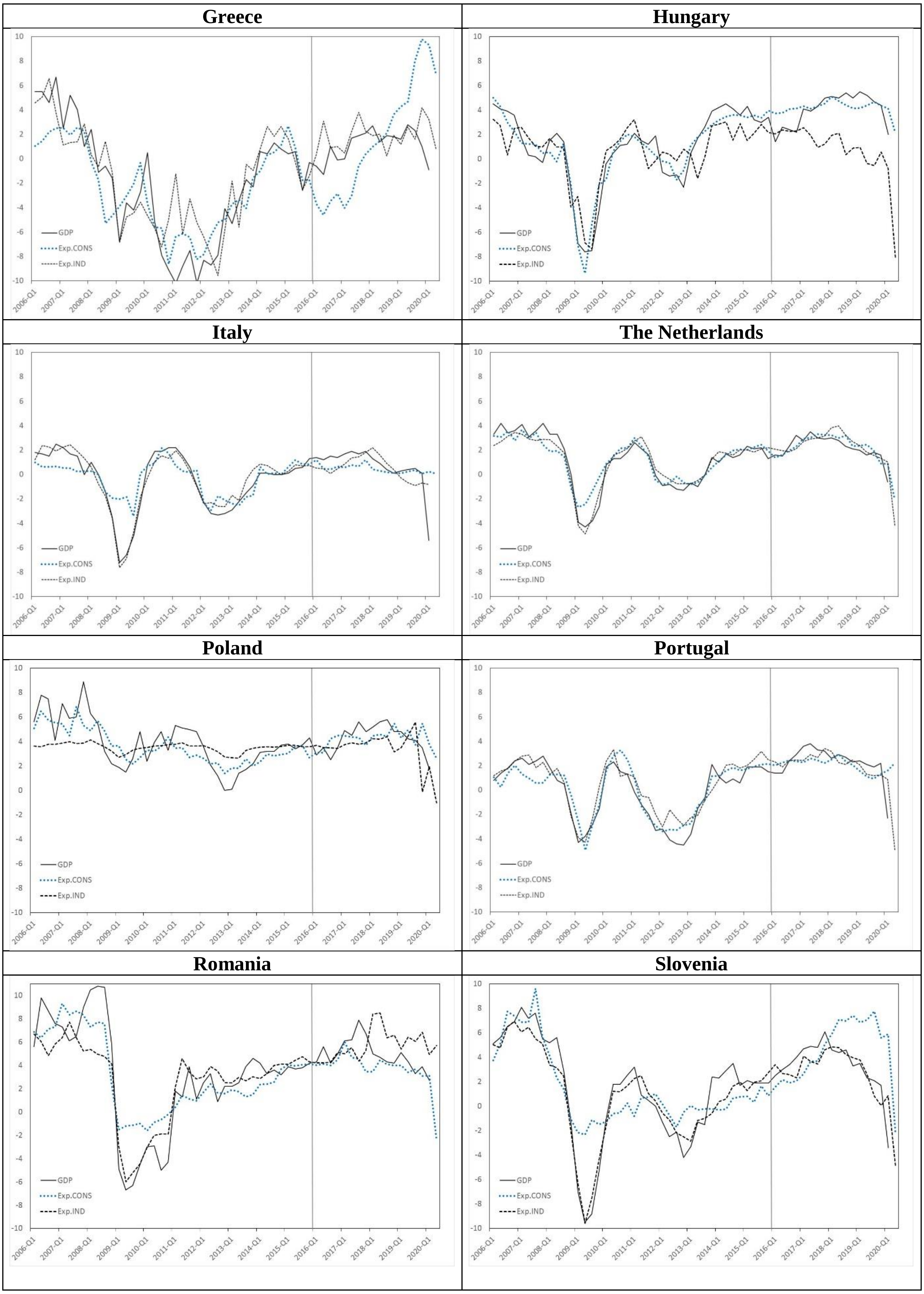

Fig. 2. (cont.1) Evolution of GDP and firms' and consumers' evolved confidence indicators
Notes: The black line represents the evolution of GDP growth, the grey dotted line the evolution of consumer confidence (Exp.CONS), and the dashed black line the evolution of industrial confidence (Exp.IND). The vertical line in 2016.Q1 marks the beginning of the out-of-sample period.

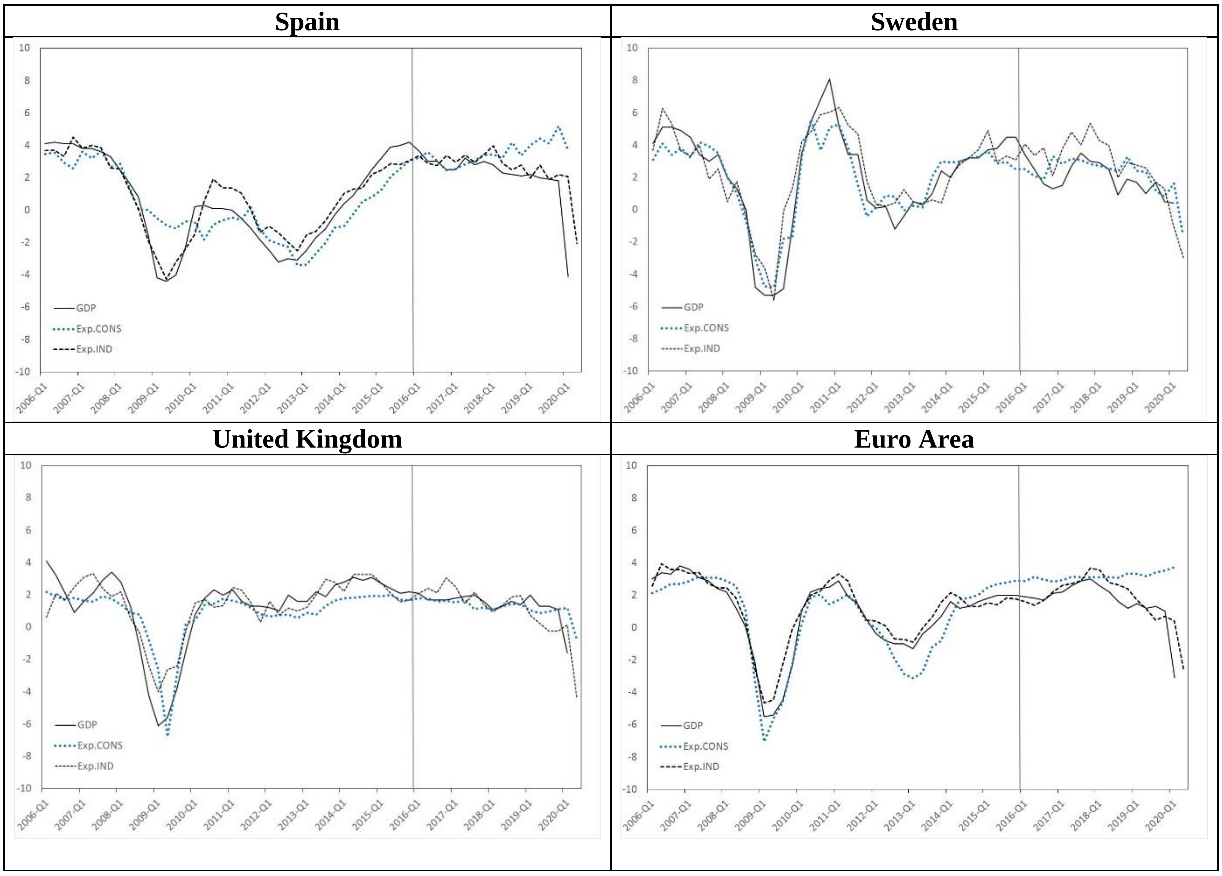

Fig. 2. (cont.2) Evolution of GDP and firms' and consumers' evolved confidence indicators
Notes: The black line represents the evolution of GDP growth, the grey dotted line the evolution of consumer confidence (Exp.CONS), and the dashed black line the evolution of industrial confidence (Exp.IND). The vertical line in 2016.Q1 marks the beginning of the out-of-sample period.

The EC publishes one composite indicator for the industry (ICI) and another one for households (CCI). Both indicators are obtained from the arithmetic mean of the balance series of a subset of questions:

$$ICI_t = \frac{I2_t + I5_t - II4_t}{3} \tag{1}$$

$$CCI_t = \frac{C2_t + C4_t + C11_t - C7_t}{4} \tag{2}$$

The in-sample OLS estimates of the weights of each of the components of the confidence indicators published by the EC allow us computing scaled confidence indicators that are directly comparable with the evolved confidence indicators. This experiment can be regarded as a nowcasting exercise, given that for each period the indicators provide an estimation of the current state of the economy before the official figures are released, making exclusive use of the latest survey data published by the EC. For further discussion of nowcasting, see Caruso (2018) and Giannone et al. (2008), and the references cited therein.

To test whether the reduction in accuracy is statistically significant, we computed the Harvey-Leybourne-Newbold statistic (Harvey, Leybourne and Newbold, 1997), which is a modification for small samples of the Diebold-Mariano (DM) statistic (Diebold and Mariano, 1995). Under the null hypothesis that there is no significant difference in precision, the statistic follows a Student-*t* distribution. A negative sign indicates that the second model has larger forecast errors. Results are presented in Table 3.

Table 3
Forecast accuracy – RMSFE – Evolved confidence indicators vs. scaled confidence indicators

	Industry			Consumers		
	Exp.IND	Cof.IND	HLN	Exp.CONS	Cof.CONS	HLN
Austria	1.097	1.186	-0.324	1.383	1.651	-0.839
Belgium	1.089	0.954	-0.122	0.946	1.698	-4.978
Bulgaria	0.665	1.180	-2.611	0.851	0.918	-1.206
Czech Republic	1.393	1.639	-0.410	3.056	3.433	-0.392
Denmark	1.643	1.516	0.272	1.189	1.084	0.686
Finland	2.284	2.243	1.100	2.335	2.759	-1.391
France	1.633	1.461	1.197	1.737	1.711	-1.137
Germany	1.244	1.808	-2.447	0.991	2.056	-2.120
Greece	1.838	1.757	0.334	4.386	4.248	0.799
Hungary	3.865	1.229	6.700	0.838	3.656	-6.518
Italy	1.373	1.167	0.711	1.522	1.568	-3.436
Netherlands	0.706	1.040	-2.311	0.592	2.265	-7.009
Poland	1.532	2.019	-2.469	1.116	2.486	-3.235
Portugal	1.009	1.113	-1.758	1.216	1.309	0.552
Romania	2.454	3.017	-1.349	1.506	1.268	1.325
Slovenia	1.505	1.355	0.425	3.989	2.203	2.606
Spain	1.583	1.523	-1.235	2.357	1.629	0.973
Sweden	1.431	3.254	-3.018	0.952	1.520	-4.806
United Kingdom	0.895	1.232	-1.243	0.775	2.137	-6.045
Euro Area	1.025	1.093	-1.382	2.147	2.031	-0.329

Notes: HLN denotes the Harvey-Leybourne-Newbold test statistic.

In Table 3 we can observe that in most countries the lowest forecast errors are obtained using the evolved indicators (Table 1 and Table 2), although the difference in accuracy is not always statistically significant. For industry, we obtained significantly lower forecast errors for Bulgaria, Germany, the Netherlands, Poland and Sweden; however, for consumers, in Belgium, Germany, Hungary, Italy, the Netherlands, Poland, Sweden, and the UK. We also observed notable differences in accuracy between firms and households in countries like the Czech Republic and Greece.

The EC weights the confidence indicators of the surveys in order to compute the Economic Sentiment Indicator (ESI). Gelper and Croux (2010), and more recently Lukac and Cizmesija (2021), have shown that letting the aggregation weights of each component of the ESI be data-driven improves its forecasting performance. Hence, we next combined the estimations obtained from the evolved industry and consumer confidence indicators by means of constrained optimisation. We used a generalised reduced gradient non-linear algorithm to minimise the summation of squared forecast errors and imposed two restrictions: (a) the sum of both weights must equal one, and (b) the weights must be equal to or larger than zero. The resulting weights are annexed in the Appendix (Table A2).

We applied the computed relative weights to combine firms' and consumers' expectations obtained from the evolved confidence indicators (Exp.Agg) and the scaled confidence indicators (Cof.Agg). We additionally computed Cof.Agg* as the average between the expectations obtained from the scaled confidence indicators. Results of the forecasting comparison are presented in Table 4.

Table 4
Forecast accuracy – Aggregate expectations – Exp.Agg vs. Cof.Agg

	RMSFE			HLN	
	Exp.Agg	Cof.Agg	Cof.Agg*	Exp.Agg vs. Cof.Agg	Exp.Agg. vs. Cof.Agg*
Austria	1.082	1.181	1.286	-0.370	-0.501
Belgium	0.994	0.937	1.085	-0.908	-3.151
Bulgaria	0.630	0.686	0.708	-0.253	-0.598
Czech Republic	1.489	1.580	1.910	0.046	-1.007
Denmark	1.147	1.120	1.235	0.463	-0.865
Finland	2.087	2.241	2.366	0.196	-0.600
France	1.648	1.448	1.485	0.576	0.204
Germany	0.943	1.653	1.655	-2.892	-3.219
Greece	2.472	2.733	2.839	-1.211	-1.734
Hungary	0.870	3.547	2.237	-6.073	-4.721
Italy	1.365	1.168	1.258	1.034	0.085
Netherlands	0.658	1.271	1.599	-3.196	-4.817
Poland	0.932	2.226	2.197	-5.065	-5.520
Portugal	1.099	0.988	0.978	1.419	1.393
Romania	1.517	1.884	1.921	-1.260	-1.612
Slovenia	1.505	2.203	1.408	-0.247	0.391
Spain	1.677	1.493	1.494	0.139	0.010
Sweden	0.930	1.955	2.230	-4.150	-3.917
United Kingdom	0.747	1.425	1.467	-3.443	-3.912
Euro Area	1.238	0.804	0.981	0.598	-0.215

Notes: Cof.Agg* denotes the average of the the scaled confidence indicators for firms (Cof.IND) and consumers (Cof.CONS). HLN denotes the Harvey-Leybourne-Newbold test statistic.

Again, we can observe that in most cases the lowest forecast errors are obtained with the aggregated expectations coming from the proposed confidence indicators (Exp.Agg), although the difference in accuracy is only statistically significant in seven countries

(Belgium, Germany, Hungary, the Netherlands, Poland, Sweden, and the UK). We also found that data-driven weights improved the forecasting performance of the scaled confidence indicators.

This forecasting exercise addresses the question about the information content of business and consumer survey expectations, and whether more sophisticated aggregation schemes based on machine learning could provide composite indicators that can better track economic activity. Our findings are in line with recent research by Ardia et al. (2019), who found that the use of optimised news-based sentiment values yielded accuracy gains for forecasting US industrial production. For Switzerland and Germany, Iselin and Siliverstovs (2016) obtained improvements in accuracy of one-step-ahead GDP forecasts by augmenting benchmark autoregressive models with variations of the recession-word index. Similarly, Juhro and Iyke (2020) found that accounting for consumer and business sentiments led to improved forecast accuracy of consumption in Indonesia.

There is ample evidence that survey expectations are useful for predicting economic variables (Altug and Çakmakli, 2016; Claveria, 2020, 2021; Girardi et al., 2015; Klein and Özmucur, 2010; Martinsen et al., 2014). In this sense, the obtained results are consistent with recent research regarding the predictive content of survey expectations. Cepni et al. (2019) showed the usefulness of diffusion indexes from the Markit survey in nowcasting and forecasting GDP in emerging markets by means of machine-learning and dimensionality-reduction techniques. Using qualitative survey responses from the ifo's World Economic Survey (WES), Hutson et al. (2014) found that respondents provided statistically significant directional forecasts. Driver and Meade (2019) used survey data from South Africa to investigate the accuracy of directional and point forecasts of investment, and found that for shorter horizons survey forecasts enhanced by time-series data significantly improved point forecasting accuracy.

5. Iterative forecasting experiment

In order to further explore the potential of the proposed approach for short-term economic forecasting, we designed an iterative out-of-sample forecasting experiment in which we re-ran the evolutionary process for each period of the out-of-sample subset using a rolling estimation window. We compared the obtained results with autoregressive moving average (ARIMA) forecasts used as a benchmark.

In order to determine the number of lags that should be included in the model, we have selected the model with the lowest value of the Akaike Information Criterion (AIC) considering models with a minimum number of 1 lag up to a maximum of 4, including all the intermediate lags. In Table 5, we present the results of comparing the out-of-sample forecasting performance of the proposed approach to rolling ARIMA forecasts used as a benchmark for two different forecast horizons (*h*).

Table 5
Forecast accuracy – RMSFE – Iterative aggregate expectations vs. ARIMA forecasts

	h=1			*h=4*		
	SR	ARIMA	HLN	SR	ARIMA	HLN
Austria	0.474	0.883	-0.835	0.475	1.590	-2.103
Belgium	0.421	1.093	-2.729	0.555	1.139	-2.384
Bulgaria	0.650	0.316	2.190	0.928	0.688	0.668
Czech Republic	0.629	1.264	-1.493	0.929	2.578	-4.068
Denmark	0.599	1.410	-3.084	0.624	2.195	-4.073
Finland	0.657	0.832	-1.413	0.719	1.181	-1.193
France	0.908	1.541	-1.108	1.047	2.275	-1.599
Germany	0.530	0.905	-1.203	0.711	1.594	-2.581
Greece	0.523	0.952	-2.280	0.535	1.827	-3.175
Hungary	0.528	2.866	-4.146	1.829	2.845	-0.860
Italy	0.649	1.493	-1.447	1.055	1.849	-2.957
Netherlands	0.287	0.734	-1.773	1.107	1.182	0.402
Poland	0.419	2.774	-5.990	0.842	2.743	-2.458
Portugal	0.643	1.335	-1.157	1.502	1.802	-0.380
Romania	0.543	3.254	-8.019	2.326	3.329	-0.775
Slovenia	0.543	2.380	-5.018	2.048	2.574	-0.924
Spain	0.711	1.630	-0.929	0.685	1.862	-1.952
Sweden	0.274	0.955	-4.316	0.572	1.804	-6.157
United Kingdom	0.383	0.861	-2.978	0.609	1.465	-1.819
Euro Area	0.426	1.102	-0.867	0.853	1.705	-1.610

Notes: *h* denotes the forecasting horizon. SR denotes the iterative forecasts obtained with the proposed GP-based approach, and ARIMA refer to the iterative ARIMA forecasts. HLN denotes the Harvey-Leybourne-Newbold test statistic.

We find that in all countries except Bulgaria, iterative sentiment indicators produce lower RMSFE values than ARIMA models, regardless of the forecast horizon. This gain in forecast accuracy is significant in ten of the countries for one-quarter-ahead predictions (*h*=1), and in nine economies for four-quarter-ahead forecasts (*h*=4). Consequently, the iterative approach allows to refine the predictive capacity obtained in the nowcasting exercise (Table 3 and Table 4). Compared to ARIMA predictions, the relative improvement of the proposed methodology increases along with the predictive horizon. Proof of this is that the RMSFE obtained for one- and four-quarter-ahead predictions is practically identical in most countries. The explanation lies fundamentally in the fact that the generated indicators tend to show a stable behaviour over long periods.

These results show the predictive potential of the proposed procedure, and provide evidence regarding the ability of GP to solve optimisation problems related to economic modelling and forecasting. In this sense, our study connects with previous research by Chen et al. (2010), who incorporated GP in a vector error correction framework and obtained better forecasts of US imports than with ARIMA models. Using information from the ifo's WES, Claveria et al. (2017b) implemented GP to construct a perceptions index and an expectations index, obtaining more accurate forecasts with the former. Similarly, Duda and Szydło (2011) applied GP to develop a set of empirical models to forecast GDP, investment and loan rates in Poland, and found that the proposed approach outperformed artificial neural network models. Focusing on the EA, Kapetanios et al. (2016) showed the usefulness of genetic algorithms to forecast quarterly GDP growth and monthly inflation. Previous applications of evolutionary computing in finance have also shown the potential of GP for the prediction of the financial failure of firms (Acosta-González and Fernández, 2014), to forecasting exchange rates (Álvarez-Díaz and Álvarez, 2003, 2005), and for stock price forecasting (Kaboudan, 2000; Larkin and Ryan, 2008; Wilson and Banzhaf, 2009).

6. Conclusion

Economic sentiment indicators are key for monitoring the current state of the economy and providing forward-looking information regarding imminent economic developments. In this paper, we propose a machine-learning method for sentiment indicators construction. The proposed approach allows us to find optimal combinations of a wide range of qualitative survey expectations that minimise a loss function and generate quantitative estimates of economic growth. By means of genetic algorithms, we obtained country-specific industry and consumer confidence indicators that allow monitoring the dynamics of economic activity in fifteen European countries and the EA, both from the demand and the supply sides of the economy.

The obtained evolved expressions differ from the confidence indicators constructed by the EC in several ways. On the one hand, they are based on information coming from all the available variables of the industry and consumer surveys. On the other hand, they generate direct estimates of economic growth. Additionally, the proposed approach automatically selects the expectational variables with the highest forecasting power and their optimal lag structure, detecting and modelling the existing non-linear relationships between survey expectations.

An examination of the obtained mathematical expressions gives insight into the relative predictive power of each of the survey variables of the industry and the consumer surveys, and also into the optimal number of lags to be taken from each of the variables to best track year-on-year GDP growth in each country. We find that firms' production expectations for the months ahead and and consumers' assessments about the general economic situation over the previous months are, respectively, the survey variables that most frequently appear in the evolved indicators, both lagged and contemporaneous. We also observed that all questions of the consumer survey appeared in the indicators, while in the case of the industry survey the distribution between variables is less uniform, with the two questions related to production being the most frequent. These findings can be very useful when using data from business and consumer surveys for economic analysis.

Finally, we assessed the forecasting performance of the proposed indicators. On the one hand, we compared them to the confidence indicators constructed by the EC in a nowcasting exercise and found that the evolved expressions outperform the scaled confidence indicators in most cases, although the differences are only significant in less than half of the countries. On the other hand, we designed a recursive out-of-sample forecasting experiment in which we iteratively re-computed the indicators to track economic growth. The obtained predictions were then compared to recursive autoregressive moving average forecasts of GDP used as a benchmark. We found that the proposed approach significantly outperforms univariate time series models in terms of accuracy.

The obtained results provide evidence regarding the ability of genetic programming to solve optimisation problems related to economic modelling, and show the potential of the methodology as a predictive tool. Furthermore, the proposed indicators are easy to implement and help to monitor the evolution of the economy, from both the demand and the supply sides. From an economic policy point of view, we have provided managers and researchers with a set of country-specific indicators that transform the qualitative expectations of firms and consumers into advanced estimates of national GDP growth without making any assumptions regarding economic agents' behaviours.

We want to note that due to the empirical nature of the proposed approach, the evolved expressions lack any theoretical background. In this sense, an issue left for further research is the introduction of restrictions in the design of the experiments with the objective of generating expressions that admit an economic interpretation. Another limitation of the proposed approach is that, as opposed to standard regression, the significance of the parameters obtained in symbolic regression cannot be assessed. Other aspects that remain to be explored are the implementation of the analysis using mixed frequency data, as well as the extension of the analysis to other economic tendency surveys, such as the construction and retail trade surveys of the Joint Harmonised Programme of Business and Consumer Surveys conducted by the EC or the Consumer Survey of the University of Michigan.

References

Álvarez-Díaz, M. (2020). Is it possible to accurately forecast the evolution of Brent crude oil prices? An answer based on parametric and nonparametric forecasting methods. *Empirical Economics*, 10(6), 1285–1305.

Álvarez-Díaz, M., & Álvarez, A. (2003). Forecasting exchange rates using genetic algorithms. *Applied Economics Letters*, 10(6), 319–322.

Álvarez-Díaz, M., & Álvarez, A. (2005). Genetic multi-model composite forecast for non-linear prediction of exchange rates. *Empirical Economics*, 30(3), 643–663.

Ardia, D., Bluteau, K., & Boudt, K. (2019). Questioning the news about economic growth: Sparse forecasting using thousands of news-based sentiment values. *International Journal of Forecasting*, 35(4), 1370–1386.

Caruso, A. (2018). Nowcasting with the help of foreign indicators: The case of Mexico. *Economic Modelling*, 69, 160–168.

Cepni, O., Güney, I. E., & Swanson, N. R. (2019). Nowcasting and forecasting GDP in emerging markets using global financial and macroeconomic diffusion indexes. *International Journal of Forecasting*, 35(2), 555–572.

Chen, X., Pang, Y., & Zheng, G. (2010). Macroeconomic forecasting using GP based vector error correction model. In J. Wang (Ed.), *Business Intelligence in Economic Forecasting: Technologies and Techniques* (pp. 1–15). Hershey, PA: IGI Global.

Claveria, O. (2020). Uncertainty indicators based on expectations of business and consumer surveys. *Empirica*. Forthcoming.

Claveria, O. (2021). On the aggregation of survey-based economic uncertainty indicators between different agents and across variables. *Journal of Business Cycle Research*. Forthcoming.

Claveria, O., Monte, E., & Torra, S. (2017a). A new approach for the quantification of qualitative measures of economic expectations. *Quality & Quantity*, 51(6), 2685–2706.

Claveria, O., Monte, E., & Torra, S. (2017b). Using survey data to forecast real activity with evolutionary algorithms. A cross-country analysis. *Journal of Applied Economics*, 20(2), 329–349.

Claveria, O., Monte, E., & Torra, S. (2019). Evolutionary computation for macroeconomic forecasting. *Computational Economics*, 53(2), 833–849.

Claveria, O., Monte, E., Torra, S. (2020). Economic forecasting with evolved confidence indicators. *Economic Modelling*, 93, 576–585.

Claveria, O., Monte, E., & Torra, S. (2021). A genetic programming approach for estimating economic sentiment in the Baltic countries and the European Union. *Technological and Economic Development of Economy*. 27(1), 262–279.

Claveria, O., Pons, E., & Ramos, R. (2007). Business and consumer expectations and macroeconomic forecasts. *International Journal of Forecasting*, 23(1), 47–69.

Diebold, F. X., & Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3), 253–263.

Driver, C., & Meade, N. (2019). Enhancing survey-based investment forecasts. *Journal of Forecasting*, 38(3), 236–255.

European Commission (2020). *Business and consumer surveys*. Available at: https://ec.europa.eu/info/business-economy-euro/indicators-statistics/economic-databases/business-and-consumer-surveys_en.

Eurostat (2020). *Database*. Available at: https://ec.europa.eu/eurostat/data/database.

Fortin, F. A., De Rainville, F. M., Gardner, M. A., Parizeau, M., & Gagné, C. (2012). DEAP: Evolutionary algorithms made easy. *Journal of Machine Learning Research*, 13(1), 2171–2175.

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4), 665–676.

Girardi, A., Gayer, C., & Reuter, A. (2015). The role of survey data in nowcasting euro area GDP growth. *Journal of Forecasting*, 35(5), 400–418.

Harding, S., Leitner, J., & Schmidhuber, J. (2013) Cartesian genetic programming for image processing. In: Riolo, R. et al. (Eds.), Genetic Programming Theory and Practice X. Genetic and Evolutionary Computation. New York, NY: Springer.

Harvey, D. I., Leybourne, S. J., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of forecasting*, 13(2), 281–291.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference and prediction* (2nd Edition). New York: Springer Series in Statistics.

Hutson, M., Joutz, F., & Stekler, H. (2014). Interpreting and evaluating CESIfo's World Economic Survey directional forecasts. *Economic Modelling*, 38, 6–11.

Iselin, D., & Siliverstovs, B. (2016). Using newspapers for tracking the business cycle: a comparative study for Germany and Switzerland. *Applied Economics*, 48(12), 1103–1118.

Juhro, S. M., & Iyke, B. N. (2020). Consumer confidence and consumption in Indonesia. *Economic Modelling*, 89, 367–377.

Kaboudan, M. A. (2000). Genetic programing prediction of stock prices. *Computational Economics*, 16(3), 207–236.

Kapetanios, G., Marcellino, M., & Papailias, F. (2016). Forecasting inflation and GDP growth using heuristic optimisation of information criteria and variable reduction methods. *Computational Statistics and Data Analysis*, 100, 369–382.

Klein, L. R., & Özmucur, S. (2010). The use of consumer and business surveys in forecasting. *Economic Modelling*, 27(6), 1453–1462.

Koza, J. R. (1992). *Genetic programming: On the programming of computers by means of natural selection*. Cambridge, MA: MIT Press.

Larkin, F., & Ryan, C. (2008). Good news: Using news feeds with genetic programming to predict stock prices. In M. O'Neil et al. (Eds.), *Genetic Programming* (pp. 49–60). Berlin: Springer-Verlag.

Lukac, Z., & Cizmesija, M. (2014). (Re)Constructing the European Economic Sentiment Indicator: An optimization approach. *Social Indicators Research*. Forthcoming.

Martinsen, K., Ravazzolo, F., & Wulfsberg, F. (2014). Forecasting macroeconomic variables using disaggregate survey data. *International Journal of Forecasting*, 30(1), 65–77.

Nicolau, M., & Agapitos, A. (2020). Choosing function sets with better generalisation performance for symbolic regression models. *Genetic Programming and Evolvable Machines*. Forthcoming.

Sorić, P., Lolić, I., & Čižmešija, M. (2016). European economic sentiment indicator: an empirical reappraisal. *Quality & Quantity*, 50(5), 2025–2054.

Vanneschi, L., & Poli, R. (2012) Genetic programming – Introduction, applications, theory and open issues. In: Rozenberg, G. et al. (Eds.), *Handbook of Natural Computing*. Berlin: Springer.
Wilson, G., & Banzhaf, W. (2009). Prediction of interday stock prices using developmental and linear genetic programming. In M. Giacobini et al. (Eds.), *Applications of Evolutionary Computing* (pp. 172–181). Berlin: Springer-Verlag.

APPENDIX

Monthly and quarterly survey indicators from the Joint Harmonised EU Industry and Consumer surveys:

Table A1
Survey indicators

Industry survey
Monthly questions
B1 – Production trend observed in recent months
B2 – Assessment of order-book levels
B3 – Assessment of export order-book levels
B4 – Assessment of stocks of finished products
B5 – Production expectations for the months ahead
B6 – Selling price expectations for the months ahead
B7 – Employment expectations for the months ahead
Quarterly questions
B8 – Assessment of current production capacity
B9 – New orders in recent months
B10 – Export expectations for the months ahead
B11 – Current level of capacity utilization (%)
B12 – Competitive position domestic market
B13 – Competitive position inside EU
B14 – Competitive position outside EU
Consumer survey
Monthly questions
C1 – Financial situation over last 12 months
C2 – Financial situation over next 12 months
C3 – General economic situation over last 12 months
C4 – General economic situation over next 12 months
C5 – Price trends over last 12 months
C6 – Price trends over next 12 months
C7 – Unemployment expectations over next 12 months
C8 – Major purchases at present
C9 – Major purchases over next 12 months
C10 – Savings at present
C11 – Savings over next 12 months
C12 – Statement on financial situation of household
Quarterly questions
C13 – Intention to buy a car within the next 12 months
C14 – Purchase or build a home within the next 12 months
C15 – Home improvements over the next 12 months

The resulting optimal weights of both evolved indicators for each country are reported in Table A2.

Table A2
Relative weights of evolved expectations

	Firms' expectations	Consumers' expectations		Firms' expectations	Consumers' expectations
Austria	0.948	0.052	Italy	0.824	0.176
Belgium	0.727	0.273	Netherlands	0.773	0.227
Bulgaria	0.389	0.611	Poland	0.441	0.559
Czech Republic	0.675	0.325	Portugal	0.464	0.536
Denmark	0.182	0.818	Romania	0.481	0.519
Finland	0.759	0.241	Slovenia	0.000	1.000
France	0.698	0.302	Spain	0.815	0.185
Germany	0.730	0.270	Sweden	0.343	0.657
Greece	0.541	0.459	UK	0.542	0.458
Hungary	0.037	0.963	EA	0.712	0.288

Notes: Relative weights computed with a generalized reduced gradient non-linear algorithm.

While in most countries the obtained relative weight of the evolved industry confidence indicator is higher than that of the evolved consumer confidence indicator, there are several exceptions: in Bulgaria, Denmark, Hungary, Slovenia and Sweden, consumers' expectations clearly outweigh firms' expectations. In Greece, Poland, Portugal, Romania and the UK, the algorithm yields a similar weight to both expectations. These results suggest that arbitrarily chosen weights of partial confidence indicators for the construction of sentiment indexes may not necessarily result in the best predictors of economic activity (Sorić et al., 2016).

UBIREA

Institut de Recerca en Economia Aplicada Regional i Públic
Research Institute of Applied Economics

WEBSITE: www.ub-irea.com • **CONTACT:** irea@ub.edu

AQR

Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group

WEBSITE: www.ub.edu/aqr/ • **CONTACT:** aqr@ub.edu

Universitat de Barcelona
Av. Diagonal, 690 • 08034 Barcelona