Economic forecasting with evolved confidence indicators

Oscar Claveria[1*], Enric Monte[2], Salvador Torra[3]

[1] AQR-IREA, University of Barcelona
[2] Department of Signal Theory and Communications, Polytechnic University of Catalunya
[3] Riskcenter-IREA, Department of Econometrics and Statistics, University of Barcelona

Abstract. We present a machine-learning method for sentiment indicators construction that allows an automated variable selection procedure. By means of genetic programming, we generate country-specific business and consumer confidence indicators for thirteen European economies. The algorithm finds non-linear combinations of qualitative survey expectations that yield estimates of the expected rate of economic growth. Firms' production expectations and consumers' expectations to spend on home improvements are the most frequently selected variables – both lagged and contemporaneous. To assess the performance of the proposed approach, we have designed an out-of-sample iterative predictive experiment. We found that forecasts generated with the evolved indicators outperform those obtained with time series models. These results 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, both from demand and supply sides.

JEL Classification: C51; C55; C63; C83; C93

Keywords: forecasting; economic growth; qualitative survey data; business and consumer expectations; symbolic regression; evolutionary algorithms; genetic programming

Acknowledgements
We would like to thank the Editor, Angus Chu, and two anonymous referees for their useful comments and suggestions. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

[*] Corresponding Author.

Highlights

- Introduction of a machine-learning method for economic sentiment construction
- Generation of country-specific business and consumer confidence indicators
- Indicators combine qualitative survey expectations to estimate economic growth
- Detection of non-linear relationships between survey expectations
- Forecasting performance of proposed indicators superior to time series models
- Results show the potential of genetic programming for economic forecasting

1. Introduction

Prediction is 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 regarding the development of economic conditions are essential for economic forecasting.

Expectations are not directly observable. Consequently, agents' expectations tend to be elicited via surveys. In relation to experimental expectations, survey expectations present several advantages: (a) they are based on the knowledge of agents that de facto operate in the market, (b) they contain information on a wide variety of economic variables, and (c) they are available prior to the publication of official data. These features make them particularly 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 13 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, 2019b). 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 introduces the methodological approach and describes the experimental setup. Section 3 describes the data. In Section 4 we present the obtained business and consumer confidence indicators. In Section 5, we assess the predictive performance of the proposed confidence indicators in two different forecasting experiments. Finally, conclusions are given in Section 6.

2. Methodology – Genetic Programming

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. Whereas standard regression analysis starts from an ex-ante specification of the model, GP focuses on finding hidden relationships between variables. To this end, the process starts with an initial population of functions, which are evolved generation after generation until the algebraic expression that best fits the data is reached.

As symbolic regression is not restricted by a predefined model, GP evolves both the structure and the parameters of the models contained in the initial population. This feature offers an overview of the most relevant interactions between the variables, and allows the identification of unknown a priori patterns. The suitability of this methodology to carry out complex modelling processes, together with its great ductility, is attracting a growing number of researchers from diverse areas. Although GP was first used as a means of evaluating the dynamics between the variables that determine the money supply (Koza, 1992), since then its use in macroeconomics has been modest (Sorić et al., 2019). 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. See Fig. 1 for a visual description of the design of the experiment.

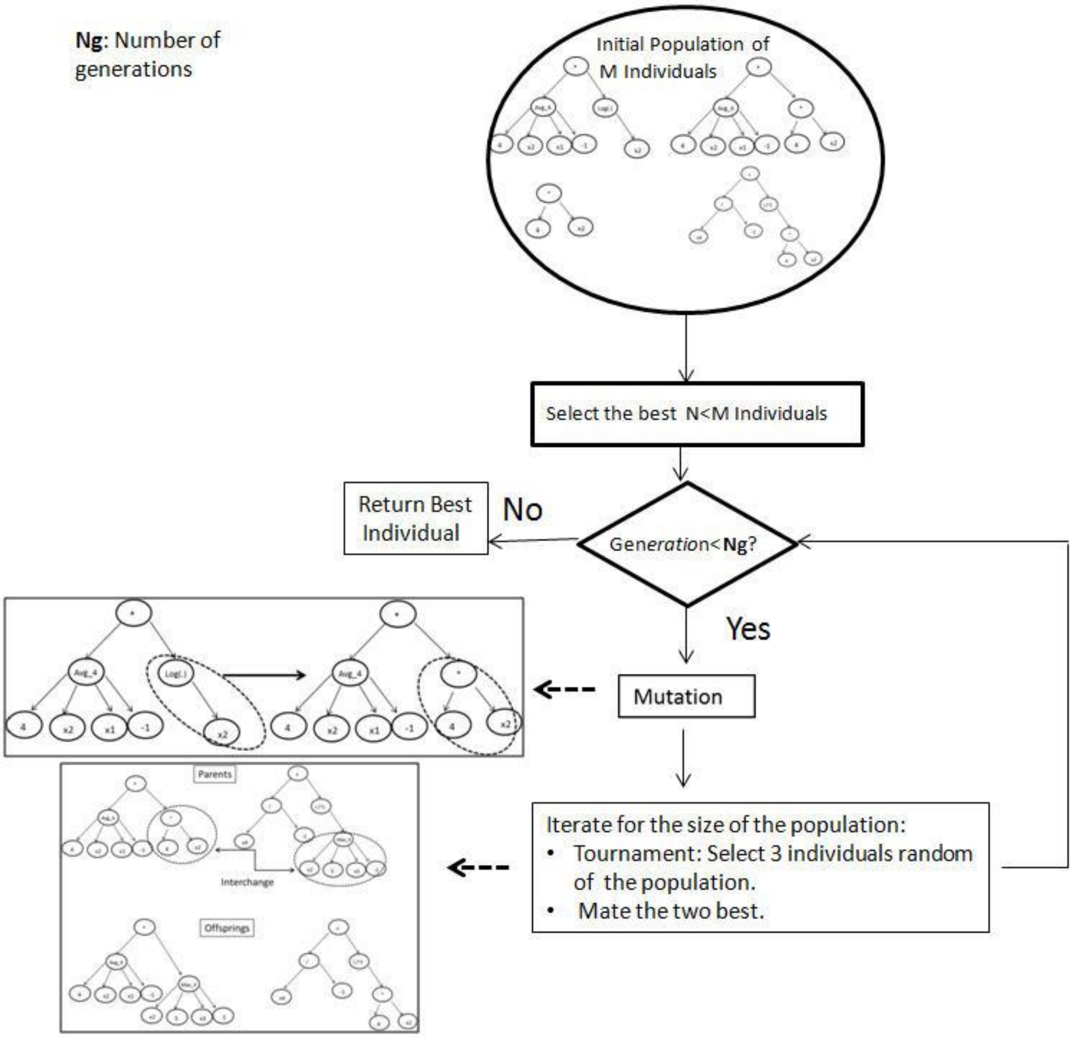

Fig. 1. Experimental setup – Design of the GP experiment

In this study, we implemented GP to generate composite indicators that capture the combinations of survey variables that best capture the dynamics of economic growth, expressed in terms of quarter-on-quarter GDP growth rates. 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. To prevent the search process of the most accurate expressions from leading to the generation of increasingly complex expressions, we have decided to limit the analysis 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 $f(x_i)$. See Hastie et al. (2009) for a justification of the need to regularise.

The term that penalises the slope is defined as the mean square of the first finite difference:

$$Reg_{slope} = \frac{1}{M-1} \sum_{i=1}^{M-1} \left(f(x_i) - f(x_{i+1}) \right)^2 \tag{1}$$

The term that penalises the curvature is defined as the mean square of the second finite difference:

$$Reg_{curvature} = \frac{1}{M-2} \sum_{i=1}^{M-2} \left(f(x_i) - 2f(x_{i+1}) + f(x_{i+2}) \right)^2 \tag{2}$$

An additional term is introduced to penalise the complexity of the solution. We denote this term as $Reg_{complexity}$. This term acts as a regularisation parameter and avoids possible overfitting due to the fact that the number of possible expressions increases very fast as the number of terms increases:

$$Reg_{complexity} = \max\left\{0\,, (Num_{operators} - 25)\right\} \tag{3}$$

One way of defining this regularisation parameter is by counting the number of terms of each expression $f(x_i)$ and, in cases when it reaches a given threshold – which in our case it has been fixed it at 25 – a penalisation proportional to the additional terms is introduced.

Finally, we define the cost function J for ranking the solutions given by the GP algorithm as a linear combination of the approximation error and the regularisation terms:

$$J = \frac{1}{M}\sum_{i=1}^{M}(f(x_i) - t_i)^2 + \alpha * Reg_{slope} + \beta * Reg_{curvature} + 0.1 * Reg_{complexity} \quad (4)$$

Coefficients (α, β) weight the regularisation terms and are selected by cross-validation in the training subset. The cross-validation is done by taking a random group of countries and computing the mean square error for a set of combinations of (α, β), and repeating the subsample process five times.

We used the Distributed Evolutionary Algorithms in Python (DEAP). The DEAP is an evolutionary computational framework built over the Python programming language, which was developed by Fortin et al. (2012) in order to facilitate the process of prototyping evolutionary algorithms due to their implementation intricacies. See Gong et al. (2015) for a comprehensive survey of state-of-the-art distributed evolutionary algorithms and models, and Vanneschi and Poli (2012) for a detailed analysis of recent challenges related to the implementation of GP.

3. Data

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 quarter-on-quarter 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 1). Monthly survey indicators were aggregated on a quarterly basis. The sample period goes from 1998.Q1 to 2020.Q1. The last 29 quarters were used as the out-of-sample period to evaluate forecast accuracy. We focused on 13 European countries – Austria (AT), Belgium (BE), Denmark (DK), Finland (FI), France (FR), Germany (DE), Greece (EL), Italy (IT), the Netherlands (NL), Portugal (PT), 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. As can be seen in Table 1, in the industry survey, manufacturers are asked about firm-specific factors, while in the consumer survey, households must respond to both subjective (micro) and objective (macro) variables.

Table 1
Survey indicators

Industry survey
Monthly questions
I1 – Production trend observed in recent months
I2 – Assessment of order-book levels
I3 – Assessment of export order-book levels
I4 – Assessment of stocks of finished products
I5 – Production expectations for the months ahead
I6 – Selling price expectations for the months ahead
I7 – Employment expectations for the months ahead
Quarterly questions
I8 – Assessment of current production capacity
I9 – New orders in recent months
I10 – Export expectations for the months ahead
I11 – Current level of capacity utilization (%)
I12 – Competitive position domestic market
I13 – Competitive position inside EU
I14 – 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

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 only exception are two variables in the industry survey (‘factors limiting production’ and ‘months of assured production’). These variables are completely different in nature because they are not expressed as balances and, as a result, they have been omitted from the analysis which is focused on the obtention of an optimal combination of balances for each survey and country.

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:

$$ICI_t = \frac{I2_t + I5_t - II4_t}{3} \tag{5}$$

$$CCI_t = \frac{C2_t + C4_t + C11_t - C7_t}{4} \tag{6}$$

Apart from theoretical considerations, the selection of variables for the construction of confidence indicators is mainly determined by their ability to track a reference series. Abberger et al. (2018) proposed a rule-based updating procedure for variable selection to improve the performance of composite indicators. With a similar purpose, Lehmann and Wohlrabe (2016) used boosting for sequential variable selection in large datasets, and investigated the evolution of the selection procedure of variables over time. While the proposed GP-based procedure for sentiment indicators construction also allows automated variable selection, it differs from the boosting algorithm in that it simultaneously updates the relationships between the selected variables, giving the same importance to all data and allowing for non-linear interactions.

4. Evolved economic sentiment indicators

In this section, we present the industry and consumer confidence indicators obtained for each country and for the EA after the evolutionary process (1998.Q1–2012.Q4), which covers up to the last in-sample period. The last 29 quarters of the sample are used as the out-of-sample period (2013.Q1–2020.Q1). The obtained industrial and consumer confidence indicators are respectively presented in Table 2 and Table 3.

Table 2
Evolved industrial confidence indicators

Austria	$\frac{I5_t * I10_t + I5_{t-4} * I9_{t-2} + I10_t * I9_{t-2} * (I1_t - 2.0 * I6_{t-1} + 2.0 * I5_t)}{I10_t * I11_{t-1} * I9_{t-2}}$
Belgium	$\frac{(I8_{t-2} + 2.0 * I5_t) * I9_t * I5_t * I4_{t-4}}{2.0 * I8_{t-2} + 2.0 * I5_t + I11_{t-4} * I9_t * I5_t * I4_{t-4}}$
Denmark	$\frac{-I1_t + I6_{t-2} + (I1_t - I6_{t-2} + I5_t) * (I10_t - I6_{t-4} + I13_{t-4})}{(I10_t - I6_{t-4} + I13_{t-4}) * (I11_{t-1} + I14_{t-1} + I14_{t-4})}$
Finland	$\frac{I5_t * (I1_{t-1} * I11_{t-2} + 6.5)}{I9_{t-2} * I11_{t-2} + (I1_{t-1} * I11_{t-2} + 6.5) * (0.3 * I11_t + I1_{t-1})}$
France	$\frac{-I11_t - I8_{t-1} + 2.0 * I5_t * I8_{t-3} * (I5_t - I8_{t-3})}{I8_{t-3} * (I5_t - I8_{t-3}) * (I6_{t-4} + I11_{t-4})}$
Germany	$\frac{-\frac{I12_t}{\frac{I6_t}{I7_{t-3}} + I7_{t-4} - 2.0} + \frac{I6_{t-2}}{I7_{t-3}} + I8_{t-2} + I5_t + I9_t}{I11_t}$
Greece	$\frac{I1_t - 2.0 * I10_{t-1} + 3.0 * I5_t - I9_{t-3} + I1_{t-4} + I7_{t-4}}{I11_t + I7_t}$
Italy	$\frac{I1_t - I11_{t-4} + (I8_{t-1} * I6_{t-2} + I4_t) * (I2_t + I14_{t-2} - I3_{t-3} + I5_t)}{I11_{t-3} * (I8_{t-1} * I6_{t-2} + I4_t)}$
Netherlands	$\frac{4.8 * I5_t * I2_{t-2}}{I2_{t-2} * (-I5_{t-2} + I11_{t-2} + I5_t) + 6.2}$
Portugal	$\frac{(I10_t + 2.0 * I5_t) * \left(I10_t + I5_t + I5_t * (I10_t + I13_{t-2})\right)}{I6_{t-1} * I5_t + \left(I10_t + I5_t + I5_t * (I10_t + I13_{t-2})\right) * (I10_t + I11_{t-4})}$
Spain	$\frac{2.0 * I5_{t-1} I10_t + I9_{t-1} + 2.0 * I5_t + I5_{t-4} + I12_{t-4} + I9_t}{I8_{t-1} + I5_{t-1} * I11_{t-2}}$
Sweden	$\frac{I7_{t-4} * \left(I11_t + I5_t + I5_t * (-I2_{t-3} + I5_t + I9_t)\right)}{I5_t * (I5_{t-3} + I7_{t-4} + I11_t * I7_{t-4})}$
UK	$\frac{I10_{t-1} + I13_{t-1} * (-I3_t - I3_{t-2} + 2.0 * I5_t)}{(I13_{t-1}) * (I11_t - I3_{t-1} - I6_{t-3} + I5_t + I13_{t-4})}$
EA	$\frac{(I3_{t-2} + 2.0 * I5_t * I7_{t-3}) * (I7_{t-3} + I8_{t-4} + 2.0)}{\left(2.0 * I1_t + 0.5 * I11_{t-4} * (I7_{t-3} + I8_{t-4} + 2)\right) * I7_{t-3}}$

Table 3
Evolved consumer confidence indicators

Austria	$\frac{(C4_{t-1} + C7_{t-4}) * C7_{t-4}}{C13_{t-3} - C13_{t-3} * C7_{t-4}}$
Belgium	$\frac{(C15_{t-2} - C9_t) * (C14_t + C15_t * C11_{t-1} + C15_{t-2}) - C14_{t-2}}{C13_{t-4} * (C14_t + C15_t * C11_{t-1} + C15_{t-2})}$
Denmark	$\frac{(C4_{t-1} + C5_{t-2}) * (C4_{t-1} + C5_{t-1} - C5_{t-2})}{C15_t + (C15_t + C1_{t-2}) * (C4_{t-1} + C5_{t-1} - C5_{t-2})}$
Finland	$\frac{C8_{t-2} * (C8_{t-2} * C8_{t-2})}{C3_t + C9_{t-4} + (C8_{t-2} * C8_{t-2}) * (C11_{t-4} - C2_t)}$
France	$\frac{C11_{t-4} * C11_t}{5.0 * C3_{t-4} + C11_t * (C8_{t-1} + C15_{t-4}) + 0.5}$
Germany	$\frac{C7_{t-4}}{\frac{C7_{t-4}}{C11_t} - C14_{t-4} - \frac{1.0}{C7_{t-4}}}$
Greece	$\frac{C3_{t-4} * (C3_{t-1} + C10_{t-2})}{C10_{t-2} + C3_{t-4} * (C11_{t-2} - C3_{t-1} + C8_{t-4}) - C3_{t-1}}$
Italy	$\frac{C4_{t-1} + C11_{t-3}}{C3_{t-1} - C6_{t-3} * C9_t}$
Netherlands	$\frac{(C15_{t-2} + C7_t) * (C9_{t-1} + C7_t + C15_{t-4})}{C5_{t-4} + (C13_t + C15_{t-4}) * (C9_{t-1} + C7_t + C15_{t-4})}$
Portugal	$\frac{C4_{t-2} - (C11_{t-1} - C4_t) * (C9_{t-3} * C13_{t-3} * C7_{t-4})}{C8_{t-3} * (C9_{t-3} * C13_{t-3} * C7_{t-4})}$
Spain	$\frac{C5_t * (C15_t - C3_t + C7_t) - C3_t - C15_{t-1}}{C14_{t-1} * C5_t}$
Sweden	$\frac{C5_{t-2} - 0.4 * C1_{t-3} + C5_t - C4_{t-1}}{C5_{t-1} - C11_t}$
UK	$\frac{-C4_{t-1}}{C15_t + C5_{t-1} + C14_{t-1} - 2.0 * C3_t + C14_{t-3}} + 0.5$
EA	$\frac{(C5_{t-2} + C15_{t-2}) * C11_{t-3}}{C12_{t-3} + C11_{t-3} * (C15_{t-1} + C5_{t-2} - C4_t + C15_{t-2})}$

We ran two independent experiments for each country. In the first one, we linked quarter-on-quarter GDP growth rates to the industry survey indicators up until 2012.Q4. 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 (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 derived expressions presented in Table 2 and Table 3 are non-linear combinations of survey variables and, include ratios and complex interactions between survey indicators. Regarding the lag structure, most variables tend to appear indistinctly with and without lags, sometimes for the same country. All survey questions appear in the indicators, although those that are selected less frequently, they mostly appear contemporaneously.

The results of Table 2 and Table 3 are summarised in Fig. 2. In the bar chart which shows the relative frequency with which each survey variable appears in the evolved expressions, we can observe that variable *I5* from the industry survey ('production expectations for the months ahead') is the most frequent of the evolved industry confidence indicators. Regarding consumer expectations, variable *C15* ('expectations of spending on home improvements over the next 12 months') is the variable most frequently selected by the algorithm, both contemporaneous and with lags.

Klein and Özmucur (2010) also found evidence of the predictive potential of variable *I5* 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 the variables contained in the consumer quarterly surveys, *C13* ('intention to buy a car'), *C14* ('purchase a home'), and especially *C15* ('spend on home improvements over the next 12 months') are frequently selected by the algorithm in the evolved consumer indicators. Notwithstanding, in spite of their leading properties, quarterly variables have always been omitted by the EC in the construction of the official consumer confidence indicators.

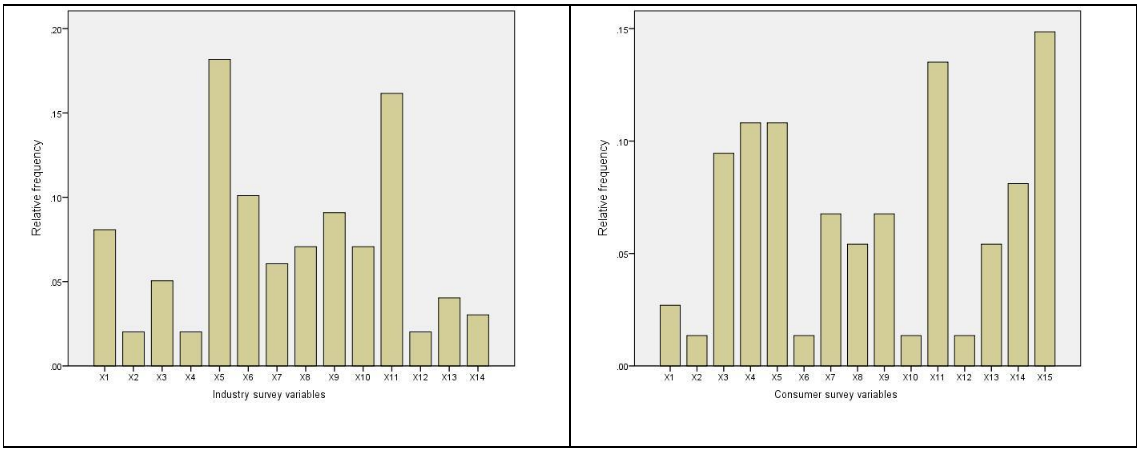

Fig. 2. Bar chart with relative frequency of variable selection (industry and consumer survey)

5. Out-of-sample forecasts

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 29 quarters (2013.Q1 to 2020.Q1) as the out-of-sample period, and the root mean square forecasting error (RMSFE) as a measure of forecast accuracy. In the first experiment, we started by using the latest available survey data to generate estimations of quarter-on-quarter GDP growth at the end of each quarter with respect to that quarter, prior to the publication of official data. In Fig. 3 we graphically compare the evolution of the two GP-generated indicators to that of the GDP of each country.

Next, we compared the out-of-sample estimations 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 quarter-on-quarter GDP growth rates, we re-scaled the indicators presented in expressions (5) and (6), by regressing the GDP growth of each country on the components of the indicators during the in-sample period (1998.Q1 to 2012.Q4).

The OLS estimates of the weights allow us to compute scaled confidence indicators that are directly comparable with the evolved confidence indicators. This experiment can be regarded as a nowcasting exercise, given that at the end of each quarter 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.

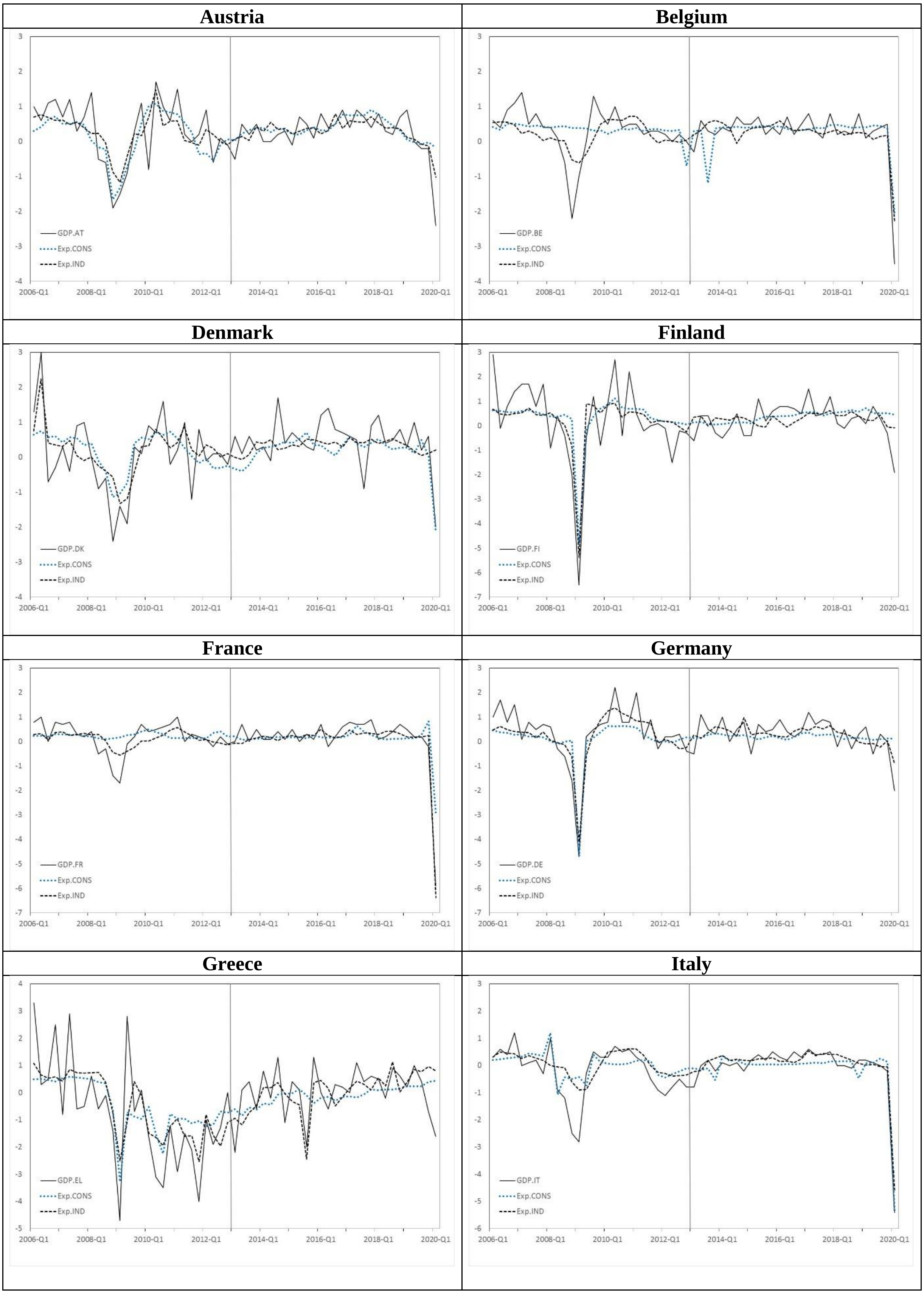

Fig. 3. Evolution of GDP and firms' and consumers' evolved confidence indicators

Notes: The black line represents the evolution of quarterly 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 2013.Q1 marks the beginning of the out-of-sample period.

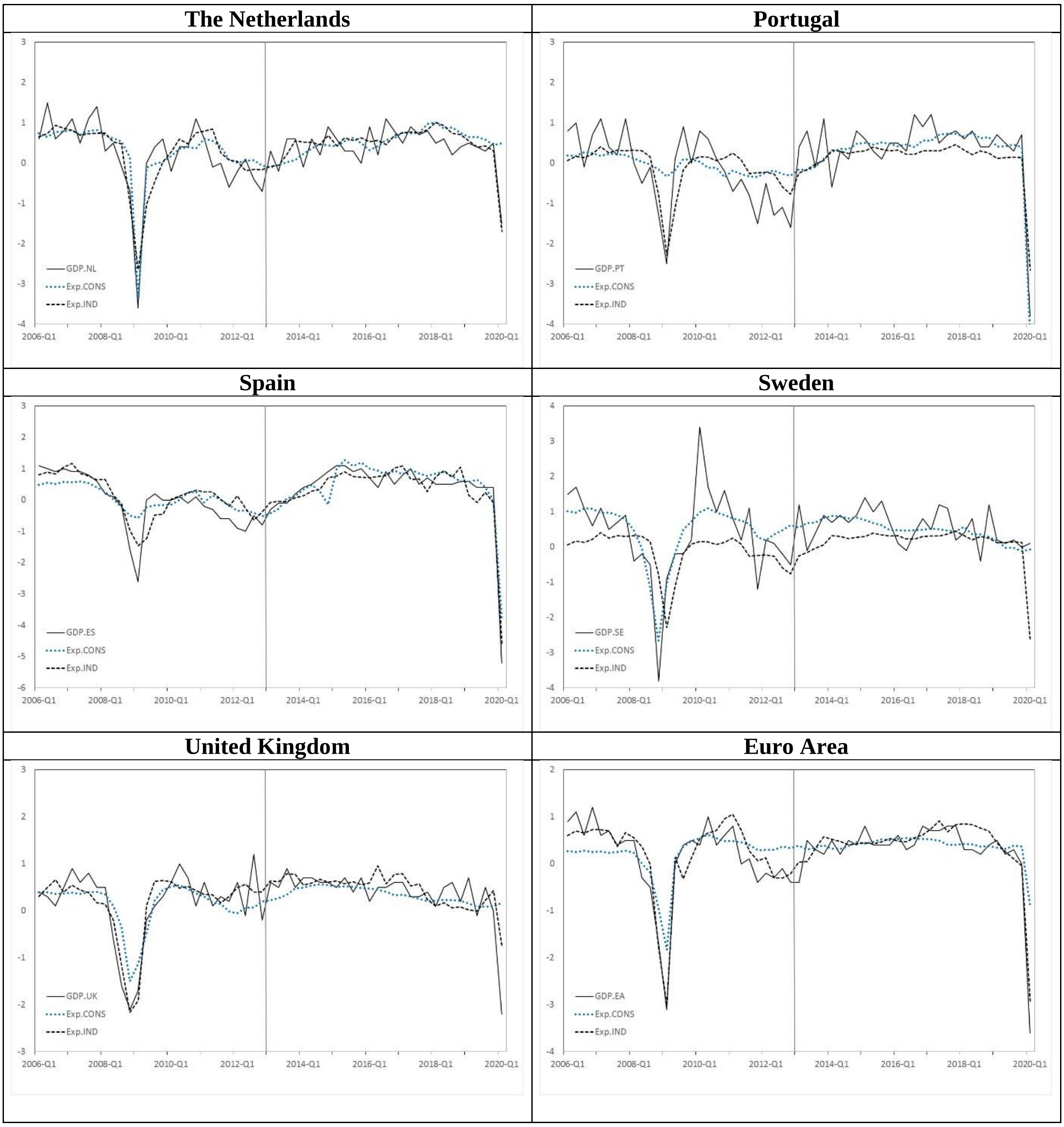

Fig. 3. (cont.1) Evolution of GDP and firms' and consumers' evolved confidence indicators

Notes: The black line represents the evolution of quarterly 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 2013.Q1 marks the beginning of the out-of-sample period.

To compare the forecast accuracy between the evolved indicators and the scaled confidence indicators, we also computed the Harvey-Leybourne-Newbold (HLN) statistic (Harvey et al., 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 standard normal distribution. A negative sign indicates that the second model has larger forecast errors. Results are presented in Table 4.

Table 4
Forecast accuracy – RMSFE – Evolved confidence indicators vs. scaled confidence indicators

	Industry			Consumers		
	Exp.IND	Cof.IND	HLN	Exp.CONS	Cof.CONS	HLN
Austria	0.426	0.778	-3.831	0.545	0.914	-2.863
Belgium	0.376	0.805	-0.630	0.360	0.826	-2.833
Denmark	0.729	1.002	-2.248	0.594	0.813	-2.594
Finland	0.628	0.917	-1.905	0.670	1.222	-2.296
France	0.292	1.246	-1.095	0.688	1.528	-9.398
Germany	0.469	0.853	-2.913	0.625	1.000	-3.849
Greece	0.799	2.023	-6.905	0.914	1.760	-2.874
Italy	0.247	1.568	-6.531	0.300	1.121	-2.342
Netherlands	0.332	0.560	-2.065	0.520	0.526	-0.456
Portugal	0.504	0.959	-1.471	0.347	0.821	-1.320
Spain	0.342	1.099	-0.671	0.425	1.082	-2.566
Sweden	0.774	1.188	-3.145	0.416	2.280	-11.893
United Kingdom	0.400	1.536	-7.515	0.521	1.231	-6.361
Euro Area	0.272	1.119	-5.415	0.572	0.854	-4.312

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

In Table 4 we observe that in all countries the lowest forecast errors are obtained using the evolved indicators presented in Table 2 and Table 3, although the difference in accuracy is not always statistically significant. For industry, we obtained significantly lower forecast errors with the evolved indicators in all countries except Belgium, Finland, France, Portugal and Spain; however, for consumers, the reduction in forecast accuracy was significant in all countries except the Netherlands and Portugal. We also found differences in accuracy between firms and households, generally obtaining lower RMSFE values with the industry evolved indicators.

The EC weights the confidence indicators of the surveys in order to compute the Economic Sentiment Indicator (ESI). Given the previous evidence that the predictive capacity of the ESI improves when the aggregation of the components is data-driven (Gelper and Croux, 2010), we next combined the estimations obtained from the evolved confidence indicators of the industry and consumers by means of constrained optimisation, minimising the summation of squared errors with a reduced gradient algorithm. The resulting optimal weights of both evolved indicators for each country are reported in Table 5.

While in most countries the obtained relative weight of the proposed industry confidence indicator is higher than that of the obtained consumer confidence indicator, there are several exceptions: in Denmark, Portugal, and Sweden, consumers' expectations outweigh firms' expectations. These results confirm 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.

Table 5
Relative weights of evolved expectations

	Firms' expectations	Consumers' expectations		Firms' expectations	Consumers' expectations
Austria	1.000	0.000	Italy	0.668	0.332
Belgium	0.733	0.267	Netherlands	0.907	0.093
Denmark	0.248	0.752	Portugal	0.123	0.877
Finland	0.682	0.318	Spain	0.718	0.282
France	0.878	0.122	Sweden	0.000	1.000
Germany	1.000	0.000	UK	1.000	0.000
Greece	0.710	0.290	EA	1.000	0.000

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

We applied the computed relative weights displayed in Table 5 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 6.

Again, we can observe that in all 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 when applying data-driven weights to combine 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 incorporating variations of the recession-word index in benchmark autoregressive models. Juhro and Iyke (2020) also found that business and consumer sentiments improved the accuracy of the Indonesian consumption forecasts.

Table 6
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	0.426	0.778	0.605	-3.831	-1.562
Belgium	0.363	0.758	0.749	-0.919	-1.277
Denmark	0.591	0.726	0.733	-1.581	-1.110
Finland	0.629	0.970	1.021	-1.926	-1.974
France	0.281	1.220	1.253	-0.985	-2.037
Germany	0.566	0.694	0.573	-3.206	-0.629
Greece	0.830	1.563	1.564	-2.187	-1.927
Italy	0.234	1.221	1.102	-2.930	-1.524
Netherlands	0.328	0.529	0.452	-1.654	-1.010
Portugal	0.342	0.821	0.854	-1.269	-1.686
Spain	0.327	1.069	1.059	-0.921	-1.236
Sweden	0.416	2.280	0.746	-11.893	-4.134
United Kingdom	0.400	1.536	0.623	-7.515	-1.482
Euro Area	0.272	1.119	0.677	-5.415	-1.087

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.

There is ample evidence that survey expectations are useful for predicting economic variables (Altug and Çakmakli, 2016; Claveria et al., 2007; 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 to forecast GDP in emerging markets by means of machine-learning and dimensionality-reduction techniques. Claveria et al. (2019a) used survey indicators from the ifo's World Economic Survey (WES) to generate expectations of economic growth in four Scandinavian economies, and found an improvement in the capacity of agents to anticipate the evolution of GDP after the 2008 crisis.Using qualitative survey responses from the same survey, Hutson et al. (2014) found that respondents provided statistically significant directional forecasts. Recently, Driver and Meade (2019) used survey data from South Africa to investigate the accuracy of investment forecasts, and found that for short horizons the combination of time-series and survey forecasts led to improvements in accuracy.

With the aim of further exploring 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 an expanding estimation window. We compared the obtained results with autoregressive moving average (ARIMA) forecasts used as a benchmark. We used Akaike's information criterion (AIC) for model selection, considering models with a minimum number of 1 lag up to a maximum of 4, including all the intermediate lags. See Brockwell and Davis (2016) for an introduction to the mathematical background to ARIMA models. In Table 7, we present the results of comparing the out-of-sample forecasting performance of the proposed approach to expanding ARIMA forecasts used as a benchmark for two different forecast horizons (*h*), one-quarter-ahead (*h*=1), and four-quarter-ahead forecasts (*h*=4).

Table 7
Forecast accuracy – RMSFE – Iterative aggregate expectations vs. ARIMA forecasts

	h=1			*h=4*		
	SR	ARIMA	HLN	SR	ARIMA	HLN
Austria	0.307	0.553	-3.630	0.289	0.733	-2.364
Belgium	0.396	0.866	-3.468	0.385	0.800	-2.366
Denmark	0.378	0.759	-3.318	0.394	0.822	-5.264
Finland	0.296	0.632	-4.996	0.315	0.734	-4.068
France	0.657	1.229	-2.281	0.619	1.323	-1.582
Germany	0.279	0.636	-6.255	0.308	0.742	-5.750
Greece	0.469	0.865	-3.659	0.494	1.318	-3.547
Italy	0.554	1.082	-1.765	0.579	1.163	-2.347
Netherlands	0.225	0.544	-4.229	0.289	0.599	-2.544
Portugal	0.457	0.953	-3.075	0.510	1.023	-3.899
Spain	0.575	1.143	-1.656	0.697	1.275	-1.465
Sweden	0.244	0.541	-3.280	0.268	0.845	-3.511
United Kingdom	0.276	0.552	-2.722	0.288	0.630	-3.843
Euro Area	0.404	0.754	-2.788	0.416	0.878	-3.354

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 found that in all countries, the iteratively-generated sentiment indicators produce lower RMSFE values than ARIMA models, regardless of the forecast horizon. This gain in forecast accuracy is significant in all but two of the countries both for *h*=1 and for *h*=4. Consequently, the iterative approach allows refining of the predictive capacity obtained in the previous nowcasting exercise (Table 4 and Table 6). 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, despite the fact that their structure morphs during the evolution, 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 demonstrated the potential of GP, both for detecting the financial failure of companies (Acosta-González and Fernández, 2014), as well as for forecasting exchange rates (Álvarez- Díaz and Álvarez, 2003, 2005), and stock prices (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. We used genetic algorithms to obtain country-specific industry and consumer confidence indicators that allow independent monitoring of the dynamics of quarter-on-quarter GDP growth in thirteen European countries and the EA.

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 economic growth in each country. We find that firms' production expectations for the months ahead and and consumers' expectations to spend on home improvements over the next 12 months are, respectively, the survey variables that most frequently appear in the evolved indicators, both lagged and contemporaneous. We also observed that consumers' expectations obtained in the quarterly survey questions were frequently selected by the algorithm. 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 significantly outperform the scaled confidence indicators in most cases, especially for the consumer indicators. 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 quarter-on-quarter GDP growth 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 quarterly GDP growth for each country without making assumptions about the behaviour of economic agents.

We want to note that due to the empirical nature of the proposed approach, the evolved expressions lack any theoretical background. The introduction of restrictions in the design of the experiments, with the objective of generating expressions that admit an economic interpretation, is an aspect left for further research. 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. In that sense, the evaluation of the stability of the generated structures through Monte Carlo simulations also remains to be explored. Other aspects that remain to be analysed 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 conducted by the EC or the Consumer Survey of the University of Michigan.

References

Abberger, K., Graff, M., Siliverstovs, B., & Sturm, J. E. (2018). Using rule-based updating procedures to improve the performance of composite indicators. *Economic Modelling*, 68, 127–144.

Acosta-González, E., & Fernández, F. (2014). Forecasting financial failure of firms via genetic algorithms. *Computational Economics*, 43(2), 133–157.

Altug, S., & Çakmakli, C. (2016). Forecasting inflation using survey expectations and target inflation: Evidence from Brazil and Turkey. *International Journal of Forecasting*, 32(1), 138–153.

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.

Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting* (3rd Edition). New York, USA: Springer.

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., 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. (2019a). Evolutionary computation for macroeconomic forecasting. *Computational Economics*, 53(2), 833–849.

Claveria, O., Monte, E., & Torra, S. (2019b). Empirical modelling of survey-based expectations for the design of economic indicators in five European regions. *Empirica*, 46(2), 205–227.

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.
Duda, J., & Szydło, S. (2011). Collective intelligence of genetic programming for macroeconomic forecasting. In P. Jędrzejowicz et al. (Eds.), *Computational Collective Intelligence. Technologies and Applications* (pp. 445–454). Berlin: Springer.
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.
Gelper, S., & Croux, C. (2010). On the construction of the European economic sentiment indicator. *Oxford Bulletin of Economics and Statistics*, 72(1), 47–62.
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.
Gong, Y. J., Chen, W. N., Zhan, Z. H., Zhang, J., Li, Y., Zhang, Q., & Li, J. J. (2015). Distributed evolutionary algorithms and their models: A survey of the stat-of-the-art. *Applied Soft Computing*, 34, 286–300.
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.
International Monetary Fund (IMF) (2020). A crisis like no other, an uncertain recovery. *World Economic Outlook*, June, 2020. Washington, DC: IMF.
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.
Lehmann, R., & Wohlrabe, K. (2016). Looking into the black box of boosting: the case of Germany. *Applied Economics Letters*, 23(17), 1229–1233.
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., Claveria, O., Monte, E., & Torra, S. (2019). Unemployment expectations: A socio-demographic analysis of the effect of news. *Labour Economics*, 60, 64–74.

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.