Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics
Document de Treball 2020/01, 29 pàg.
Working Paper 2020/01, 29 pag.

Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group
Document de Treball 2020/01, 29 pàg.
Working Paper 2020/01, 29 pag.

"Economic determinants of employment sentiment: A socio-demographic analysis for the euro area"

Oscar Claveria, Ivana Lolić, Enric Monte, Salvador Torra, Petar Sorić

UBIREA
Research Institute of Applied Economics
AQR
Regional Quantitative Analysis Research Group

UBIREA | Institut de Recerca en Economia Aplicada Regional i Pública
UNIVERSITAT DE BARCELONA

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

The Research Institute of Applied Economics (IREA) in Barcelona was founded in 2005, as a research institute in applied economics. Three consolidated research groups make up the institute: AQR, RISK and GiM, and a large number of members are involved in the Institute. IREA focuses on four priority lines of investigation: (i) the quantitative study of regional and urban economic activity and analysis of regional and local economic policies, (ii) study of public economic activity in markets, particularly in the fields of empirical evaluation of privatization, the regulation and competition in the markets of public services using state of industrial economy, (iii) risk analysis in finance and insurance, and (iv) the development of micro and macro econometrics applied for the analysis of economic activity, particularly for quantitative evaluation of public policies.

Abstract

In this study we construct quarterly consumer confidence indicators of unemployment for the euro area using as input the consumer expectations for sixteen socio-demographic groups elicited from the Joint Harmonised EU Consumer Survey. First, we use symbolic regressions to link unemployment rates to qualitative expectations about a wide range of economic variables. By means of genetic programming we obtain the combination of expectations that best tracks the evolution of unemployment for each group of consumers. Second, we test the out-of-sample forecasting performance of the evolved expressions. Third, we use a state-space model with time-varying parameters to identify the main macroeconomic drivers of unemployment confidence and to evaluate whether the strength of the interplay between variables varies across the economic cycle. We analyse the differences across groups, obtaining better forecasts for respondents comprised in the first quartile with regards to the income of the household and respondents with at least secondary education. We also find that the questions regarding expected major purchases over the next 12 months and savings at present are by far, the variables that most frequently appear in the evolved expressions, hinting at their predictive potential to track the evolution of unemployment. For the economically deprived consumers, the confidence indicator seems to evolve independently of the macroeconomy. This finding is rather consistent throughout the economic cycle, with the exception of stock market returns, which governed unemployment confidence in the pre-crisis period.

JEL Classification: C51, C53, C55, D12, E24, E27, J10

Keywords: Unemployment, Expectations, Consumer behaviour, Forecasting, Genetic programming, State-space models yield.

Oscar Claveria: AQR-IREA, University of Barcelona. Email: oclaveria@ub.edu

Ivana Lolić: University of Zagreb, Faculty of Economics and Business. Email: ilolic@efzg.hr

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

Salvador Torra: Riskcenter-IREA, Department of Econometrics and Statistics, University of Barcelona. Email: storra@ub.edu

Petar Sorić: University of Zagreb, Faculty of Economics and Business. Email: psoric@efzg.hr

Acknowledgements

This research was supported by the project HRZZ IP-2018-01-4189 form the Croatian Science Foundation

1. Introduction

Recent research has shown that in the last three decades euro area (EA) member states have suffered more frequent and more intensive recessions than other developed economies (Bluedorn et al., 2019). Putting that in the context of the recent political and economic uncertainty related to Great Britain's withdrawal from the European Union and the slowing down of major European economies, the revival of interest in developing leading indicators and forecasting the macroeconomy seems well grounded (Rossi and Sekhposyan, 2017; Perić and Sorić, 2018). Having in mind that the last crisis has triggered an unemployment hysteresis (Krištić et. al., 2019), correctly anticipating the unemployment rate becomes crucial. Since the second quarter of 2013, the unemployment rate in the EA has shown a downward trend (Eurostat, 2019), but knowing for how much longer remains key.

Unemployment forecasts are typically made using the historical time-series properties of the unemployment rate and indicators of the labour market (Barnichon et al., 2012; Claveria, 2019a,b; Hutter and Weber, 2015), or using Okun's law as the underlying model (Ball et al., 2015). In this paper we use a new approach exclusively based on the information coming from consumer economic expectations. Our motivation for building such indicators for sixteen socio-demographic groups in the EA stems from the premise that a new wave of an unemployment hysteresis might induce severe political and social instabilities and a fall of trust in the European integration process.

Having that in mind, the main aim of the paper is twofold. On the one hand, by means of genetic programming (GP) we evolve a quarterly leading indicator of unemployment confidence for each group designed to generate forecasts of the unemployment rate. Our main objective is to provide researchers and forecasters with an easy tool to generate short-term predictions of the unemployment rate. On the other hand, we examine the driving forces of unemployment confidence in the EA in a time-varying framework for each socio-demographic group.

The European Commission (EC) conducts monthly and quarterly surveys to collect information on households' spending and saving intentions. Consumers are asked about the past, present and future evolution of a wide range of variables concerning their economic situation. More specifically, they are asked whether they expect a variable to rise, to fall or to remain constant. The results are presented as balances, which are the subtraction between the percentage of respondents expecting an increase and the

percentage of respondents expecting a decrease. The EC constructs a set of widely-used confidence indicators by choosing a set of questions from each survey and weighting their respective response balances: the consumer confidence indicator, the industrial confidence indicator, and the construction confidence indicator.

In this study we develop a consumer confidence indicators of unemployment (CCIU) for each group of consumers. To construct the composite unemployment indicator based on consumer survey results, we use a heuristic based on GP. The choice of variables to be included and the mathematical form to be used is freely determined by an evolutionary algorithm which is set to minimise the forecast errors when tracking the evolution of the unemployment rate.

One of the main advantages of the proposed approach is that it gives insight so as to the survey variables that are key in predicting the evolution of unemployment. Given the empirical nature of the approach, the resulting evolved expression of the indicator is obtained without imposing any assumption. As survey expectations are available ahead of the publication of official quantitative statistics, the obtained indicator allows to generate one-quarter ahead forecasts of the unemployment rate.

After assessing its out-of-sample predictive performance, we use a state-space framework to evaluate the effect of the main macroeconomic variables on unemployment confidence. In a hypothetical scenario of a strong shock in unemployment confidence, the sole fact that agents expect a sharp deterioration of economic climate in the near future might trigger firms to lay off employees and factually increase unemployment. The concept of self-fulfilling prophecy is well established in economic theory (Azariadis, 1981), so it is vital to carefully monitor unemployment confidence and identify its main determinants. To do so, we use a time-varying model, aiming to give some insight into the dynamic interplay of the main drivers of unemployment confidence.

Both experiments are carried out for the sixteen different socio-economic groups, obtaining an indicator for each group of consumers and assessing the impacts of macroeconomic variables on the unemployment confidence for that particular collective.

The paper is organised as follows. The next section introduces the data and the methodological approach. In Section 3 we assess the out-of-sample forecasting performance of the proposed indicator. The evaluation of the macroeconomic drivers of unemployment sentiment is presented in Section 4. Finally, concluding remarks and future lines of research are drawn in Section 5.

2. Data and Methodology

This section briefly introduces the dataset and describes the employed methodology. We describe the experimental setup designed to extract consumers' unemployment expectations and to obtain the CCIU. As we aim to formalise the optimal interactions between consumers' expectations that best estimate the evolution of unemployment in the EA, we use two types of information: qualitative survey data and quantitative official statistics from 2005:Q1 to 2017:Q4. Regarding the former, we make use of seasonally adjusted balances from the consumer survey conducted by the EC (https://ec.europa.eu/info/business-economy-euro/indicators-statistics/economic-databases/business-and-consumer-surveys_en). We employ the information from all available monthly and quarterly questions (see Table 1).

Table 1
Consumer Survey – Variables

	Monthly questions
X1	Financial situation over last 12 months
X2	Financial situation over next 12 months
X3	General economic situation over last 12 months
X4	General economic situation over next 12 months
X5	Price trends over last 12 months
X6	Price trends over next 12 months
X7	Unemployment expectations over next 12 months
X8	Major purchases at present
X9	Major purchases over next 12 months
X10	Savings at present
X11	Savings over next 12 months
X12	Statement on financial situation of household
	Quarterly questions
X13	Intention to buy a car within the next 12 months
X14	Purchase or build a home within the next 12 month months
X15	Home improvements over the next 12 months

In the survey, consumers are asked about their expectations regarding a wide range of variables (Table 1), and they are faced with five reply options: "a lot better/much higher/sharp increase", "a little better", "stay the same", "a little worse", and "a lot worse/much lower/sharp decrease"). The aggregated percentages of the individual replies in each category are respectively denoted as *PP*, *P*, *E*, *M* and *MM*.

With regards to the quantitative information used as the target variable, we employ the seasonally-adjusted unemployment rates provided by the OECD, which are also freely available (https://data.oecd.org/unemp/unemployment-rate.htm).

Table 1 contains the fifteen survey variables used in the study, denoted as X_{it}, where *i* refers to each group of consumers and *t* to the time period. Survey variables can be divided in judgements about the present ($X8_{it}$, $X10_{it}$ and $X12_{it}$), perceptions about the past ($X1_{it}$, $X3_{it}$ and $X5_{it}$), and expectations about the future ($X2_{it}$, $X4_{it}$, $X6_{it}$, $X7_{it}$, $X9_{it}$, $X11_{it}$, $X13_{it}$, $X14_{it}$, and $X15_{it}$). We use all variables jointly to generate the CCIU.

To do so we use state-of-the-art evolutionary algorithms that search the space of mathematical expressions that best fit our dataset, i.e. the evolution of the unemployment rate in the EA. The proposed approach is based on a combination of symbolic regression (SR) and GP (Koza, 1992, 1995). Evolutionary computation is being increasingly applied to automated problem-solving in economics (Álvarez-Díaz, 2019; Claveria et al. 2018, 2019; Marković et al., 2017). This approach for model approximation applies Darwinian principles during an evolution process in which an initial population of computer programs are bred through generations in order to find a set of analytical functions that best fit the data.

In this study we implement SR via GP to find the dynamic relationship between the unemployment rate in the EA and a wide range of expectational variables (Table 1). The number of lags is limited to four, and for the sake of simplicity and replicability, the integration schemes are restricted to the main four mathematical operations. The algorithm is programmed to search for patterns across survey variables until it finds mathematical functional forms that can be regarded as the optimal combinations of survey variables that best fit the actual evolution of unemployment.

In a first step, for each socio-demographic group we use a SR model that links the fifteen variables and their corresponding lags up to four quarters to the evolution of the unemployment rate in the EA:

$$y_t = f\left(X1_t, X1_{t-1}, X1_{t-2}, X1_{t-3}, X1_{t-4}, \ldots, X15_t, X15_{t-1}, X15_{t-2}, X15_{t-3}, X15_{t-4}\right) \quad (1)$$

Where $X1_t,\ldots,X15_t$ are the different survey variables, and y_t is a scalar referring to the monthly rate of unemployment in the EA at time *t*.

In a second step, we use GP to estimate the model. This approach is particularly suitable in our case, as there is no a priori information regarding the interactions between the different survey variables. GP simultaneously evolves the structure and the parameters of the expressions.

We implement GP to evolve expression (1) until a stopping criterion is reached. Given the trade-off between accuracy and simplicity, we have chosen a maximum number of 100 generations as the stopping criterion. For each simulation, a first random population of 75000 functions is generated. Afterwards, the best 7500 elements are selected for the evolutionary phase, where genetic operators (crossover and mutation) are applied to raise the fitness of the population generation after generation. We use the mean squared error (MSE) as the fitness function. This process is repeated for each of the 16 socio-demographic groups contained in Table 2. See Claveria et al. (2017) for a detailed description on the implementation of GP, and Dabhi and Chaudhary (2015) and White et al. (2013) for a detailed review of the main issues of GP.

Table 2
Socio-demographic groups – Survey sub-categories

Income of the household	
1st quartile included	RE1
2nd quartile included	RE2
3rd quartile included	RE3
4th quartile included	RE4
Education of respondent	
Primary	ED1
Secondary	ED2
Further	ED3
Age of the respondent	
16-29	AG1
30-49	AG2
50-64	AG3
65+	AG4
Sex of the respondent	
Male	MAL
Female	FEM
Occupation of respondent	
Work full-time	PR8
Work part-time	PR9
Unemployed	PR0

In Table 3 we present a descriptive analysis of survey results for the sample period (2015:Q1–2017:Q4). Statistics are computed for each stratum according to income, education, age, gender, and occupation..

Table 3
Summary statistics – Survey results for subcategories of consumers (2005:Q1–2017:Q4)

		X1	***X2***	***X3***	***X4***	***X5***	***X6***	***X7***	***X8***	***X9***	***X10***	***X11***	***X12***	***X13***	***X14***	***X15***
RE1	mean	-24.05	-9.20	-35.61	-17.43	33.17	15.56	27.68	-26.58	-26.02	9.53	-37.52	-4.04	-85.76	-94.38	-75.69
	std.dev.	3.92	3.85	12.74	8.56	18.47	9.80	12.72	6.85	3.91	6.00	2.39	2.21	1.89	1.00	2.26
RE2	mean	-17.36	-6.68	-31.01	-13.91	28.40	14.42	24.55	-18.37	-23.33	15.25	-18.28	7.58	-78.98	-91.53	-65.32
	std.dev.	4.24	3.66	14.88	9.87	19.55	10.04	14.60	7.62	3.70	6.73	3.32	1.50	2.44	1.47	2.55
RE3	mean	-12.50	-4.08	-28.17	-12.23	25.09	13.36	22.55	-13.67	-21.20	17.81	-4.61	14.53	-72.48	-88.10	-55.63
	std.dev.	4.62	3.69	16.70	10.68	20.09	10.26	15.63	8.96	3.92	7.13	3.24	1.98	2.81	1.75	2.79
RE4	mean	-4.21	-0.37	-23.41	-8.22	18.68	11.76	19.27	-4.53	-15.85	22.64	15.46	27.66	-63.00	-82.12	-44.79
	std.dev.	4.69	3.91	19.13	12.22	21.57	11.53	17.11	10.21	3.44	7.61	3.67	2.27	3.41	1.82	3.85
ED1	mean	-18.72	-8.80	-33.74	-15.66	31.67	12.96	25.70	-22.50	-23.87	11.87	-25.81	5.68	-82.92	-94.09	-70.22
	std.dev.	4.10	3.46	13.25	8.88	18.97	9.96	13.28	7.65	3.65	7.21	3.25	1.75	1.97	0.84	2.35
ED2	mean	-13.59	-4.16	-29.44	-12.49	26.09	13.69	23.69	-13.75	-20.66	16.39	-8.75	12.06	-72.69	-88.86	-59.08
	std.dev.	4.29	3.69	15.54	9.89	19.33	10.21	14.55	8.24	3.62	7.10	2.84	1.84	2.95	1.58	2.31
ED3	mean	-7.96	-1.49	-24.20	-9.88	20.21	14.81	20.55	-5.51	-16.10	18.34	8.04	21.58	-67.13	-83.23	-50.55
	std.dev.	4.52	3.93	18.78	12.08	21.14	11.53	17.20	9.24	3.33	8.07	3.17	1.92	3.87	2.08	3.12

Notes: Std. Dev. denotes the standard deviation. Up until 2016:Q2 for PR8, PR9 and PR0.

Table 3 (cont.)
Summary statistics – Survey results for sub-categories of consumers (2005:Q1–2017:Q4)

		X1	***X2***	***X3***	***X4***	***X5***	***X6***	***X7***	***X8***	***X9***	***X10***	***X11***	***X12***	***X13***	***X14***	***X15***
AG1	mean	-4.53	5.91	-23.51	-7.03	21.04	10.38	19.79	-11.38	-15.43	25.86	4.67	15.77	-62.94	-79.62	-59.73
	std.dev.	4.70	3.44	14.97	8.59	18.31	10.15	13.05	7.77	3.61	5.97	3.55	1.96	4.00	2.62	3.19
AG2	mean	-11.88	-1.65	-29.01	-13.18	25.59	13.54	24.65	-15.83	-20.98	18.79	-5.04	12.22	-70.23	-84.92	-54.86
	std.dev.	4.91	4.21	16.21	10.58	19.89	10.78	14.99	8.42	3.81	5.99	3.84	2.00	2.88	1.95	3.00
AG3	mean	-17.69	-8.53	-31.61	-15.23	28.07	15.58	25.90	-15.66	-23.02	13.23	-12.16	11.47	-74.95	-91.10	-58.15
	std.dev.	4.43	3.84	16.43	10.81	20.48	10.64	15.55	8.56	3.48	8.23	2.90	1.61	2.31	0.92	3.19
AG4	mean	-18.17	-11.31	-30.80	-13.66	28.25	12.98	22.11	-16.41	-23.45	12.40	-19.73	11.61	-85.88	-95.94	-69.21
	std.dev.	4.11	3.55	14.83	9.96	20.45	10.34	14.86	9.31	3.43	7.95	3.67	1.35	1.90	0.66	3.45
MAL	mean	-12.34	-4.02	-26.68	-11.47	22.75	13.09	21.94	-11.74	-19.79	16.86	-7.28	13.86	-71.44	-87.97	-58.19
	std.dev.	4.38	3.80	16.55	10.74	20.81	10.81	15.77	8.95	3.46	7.38	3.09	1.61	2.79	1.33	2.85
FEM	mean	-15.88	-5.92	-32.49	-14.52	30.05	13.69	25.19	-18.56	-22.18	15.99	-13.66	10.40	-77.79	-89.98	-63.12
	std.dev.	4.24	3.69	14.67	9.65	19.03	10.06	13.86	8.05	3.68	6.70	3.13	1.55	2.05	1.25	2.57
PR8	mean	-9.22	-2.49	-30.15	-13.61	26.62	14.05	25.67	-14.58	-19.58	21.09	1.13	17.22	-68.39	-84.42	-54.36
	std.dev.	3.95	3.78	15.54	10.48	20.33	11.36	15.01	7.50	3.20	5.68	3.12	1.33	2.98	1.73	3.03
PR9	mean	-14.51	-3.45	-34.08	-15.74	30.27	14.46	27.48	-18.51	-22.77	20.16	-9.09	10.00	-72.45	-87.94	-56.86
	std.dev.	4.18	3.81	14.23	9.81	19.85	11.02	13.72	7.03	4.03	6.74	3.56	1.72	3.76	2.04	3.03
PR0	mean	-36.01	-4.69	-40.38	-19.12	33.48	16.75	33.94	-28.37	-29.75	13.24	-40.37	-11.81	-79.51	-91.16	-71.49
	std.dev.	3.85	3.98	12.67	8.65	19.67	11.09	12.38	6.39	4.27	5.83	3.02	2.87	3.38	2.35	3.31

Notes: Std. Dev. denotes the standard deviation. Up until 2016:Q2 for PR8, PR9 and PR0.

The second part of our empirical analysis deals with the determinants of unemployment expectations and their time consistency. The literature has widely acknowledged that unemployment rates exhibit a nonlinear behaviour over the business cycle (Parker and Rothman, 1997). They tend to increase abruptly in recessions, and decrease smoothly and slowly in the growth period. In this context, we want to analyse if the CCIU is also dependent on the economic cycle.

We add to the literature on psychologically-driven crises, which postulates that economic confidence gains significance in recessions (Baker et al., 2016; Christiansen et al., 2014; Garner, 1991; Sorić, 2018). In that case, the CCIU should contain additional information beyond that included in the main macroeconomic variables, and should evolve independently of the cycle during recessions.

To evaluate the time-varying effects of macroeconomic variables on unemployment confidence, we use a state-space model. The model is conceptualized as a system consisting on an observation equation (2) and a state equation (3):

$$UCI_t = \boldsymbol{z}'_{t-1}\boldsymbol{\beta}_t + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma_t^2) \tag{2}$$

$$\boldsymbol{\beta}_{t+1} = \boldsymbol{\beta}_t + \boldsymbol{\eta}_t \qquad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \boldsymbol{Q}) \tag{3}$$

Where $t = 2, \dots, T$, being T the sample size; UCI_t is the observation vector (2), and $\boldsymbol{\beta}_t{}' = (\beta_{t,0}\ \beta_{t,1})$ an unobserved state vector; $\boldsymbol{Q} = diag(\sigma_{\eta 1}^2, \sigma_{\eta 2}^2)$ is a diagonal covariance matrix, and $\boldsymbol{\eta}_t{}' = (\omega_{t,1}\ \omega_{t,2})$ is the error term vector. The vector $\boldsymbol{z}_t{}'$ is the regressor vector containing one of each of the analysed macroeconomic variables. We focus on Gross Domestic Product (GDP), private consumption, unemployment, the Harmonised Index of Consumer Prices (HICP) and stock prices.

Data of GDP, consumption and HICP are obtained from Eurostat, unemployment from the European Central Bank, and stock prices, measured by the Eurostoxx50 index, from Thomson Reuters. All the variables examined in this study are seasonally adjusted using TRAMO/SEATS, and they are aimed to reflect the economic constructs targeted through consumer survey questions.

It should be noted that each of the two variances in $\boldsymbol{Q}$, along with σ_t^2 for the error term, can be either positive (stochastic behaviour) or equal to zero (deterministic behaviour). As a result, eight model specifications are assessed for the estimation of equations (2) and (3) for each of the five selected covariates (private consumption, Eurostoxx50 index, HICP, GDP, and unemployment). We then select the optimal model specification for each combination of variables using the Akaike information criterion.

Upon selecting the optimal model specification for each combination of variables, we check the standardized residuals of the optimal model specifications for conventional assumptions (normality, non-correlated errors, and homoscedasticity). The diagnostic tests applied here are the Doornik and Hansen (2008) normality test, a nonparametric heteroskedasticity test (Koopman et al., 1999) and the Ljung-Box autocorrelation test of 4th order. Non-rejecting the null hypothesis in all three tests implies that the error terms are pure white noise processes. Equations (2) and (3) are estimated for the 2006:Q2–2017:Q4 period using the Kalman filter (Kalman, 1960).

This particular specification is used to pinpoint specific macroeconomic variables that drive consumers' unemployment sentiment. Moreover, the goal of the analysis is to scrutinize whether the strength of the observed relationships varies across the economic cycle. A glance at equation (2) reveals that the independent variables are entered in their first lag. The reasoning behind that is that consumer surveys are regularly conducted in the first two to three weeks of each month/quarter, while national accounts and other relevant macroeconomic data are officially published with a considerable time delay.

In the next section, both analyses are carried out for each of the 16 categories of consumers regarding the income of the household, the level of education, the age and gender of the respondents, and the occupation (Table 2). For each consumer category within individual stratification variables we apply GP and calculate a group-specific CCIU. Then we assess their out-of-sample forecasting performance and the dynamic effect of the main macroeconomic variables during the sample period.

3. Empirical results

In this section we first assess the forecasting performance of the composite indicators of unemployment obtained through the evolution of consumers' expectations. Additionally, we use the indicator to estimate a state-space model in order to evaluate the dynamic interplay between the main macroeconomic variables and unemployment confidence. Both analyses are replicated for each socio-demographic group.

This research adds to the sociodemographic analysis of unemployment. Even the early contributions to this field revealed an asymmetric effect of recessions on unemployment across various demographic groups. Using simulations, Smith et al. (1974) found that recessions had the most severe effect in the rise of unemployment for women and the

young. Similar findings are also obtained in more recent studies. Couch and Fairlie (2010) revealed that racial minorities' employment patterns are very sensitive to the overall business climate. Focusing on the same issue for different age groups, Xu and Couch (2017) found that younger workers have a higher chance of being fired than the older groups of workers. This conclusion is valid both before and after 2008, being even more pronounced in the recession period. It seems, therefore, that unemployment expectations should also be quite diverse with regards to agents' socio-demographic characteristics. Our analysis aims to pinpoint the main socio-demographic features that induce such divergence.

First, we present the empirically-modelled confidence indicators of unemployment. After evolving expression (1) during 100 generations by means of GP for each group of consumers, we obtain the functional forms contained in Table 4. These analytical expressions can be regarded as consumer confidence indicators. The CCIU generate one-quarter ahead forecasts of the unemployment rate in the EA for each socio-demographic group. To evaluate the performance of the evolved expressions, we assess their predictive accuracy in an out-of-sample forecasting experiment. We use the last two years of the sample to compute the mean absolute forecast error (MAFE) and the root mean squared forecast error (RMSFE):

$$MAFE_t = \frac{1}{n}\sum_{t=1}^{n}|e_t| \tag{4}$$

$$RMSFE_t = \sqrt{\frac{1}{n}\sum_{t=1}^{n}e_t^2} \tag{5}$$

Where e_t refers to the forecast error at time *t*.

In Fig. 2 we depict the mean relative frequency with which each survey variable appears in the evolved unemployment indicators. It shows that the expected major purchases over the next 12 months (*X9*), mostly lagged four quarters, and savings at present (*X10*), are the survey variables that most frequently appear in the evolved expressions, hinting at their predictive potential to track the evolution of unemployment. All fifteen survey variables appear, even if it is just in one expression. Variables *X9*, *X10*, *X13* and *X14* are the ones which repeatedly appear in one or more expressions with different lags.

Table 4
CCIU for each socio-demographic group

RE1	$-0.35 * X8_{t-4} + \frac{X3_{t-2}}{X8_{t-4}} - 0.05 * \frac{X8_{t-4}}{2.50 * X11_{t-3}}$
RE2	$-0.37 * X9_{t-4} + 1$
RE3	$-0.39 * X9_{t-4} + \frac{X13_{t-3}}{X15_{t-3}} + \frac{2.60}{X4_t + X5_{t-4}}$
RE4	$10 + \frac{X2_{t-4}}{\frac{0.1 * X2_{t-4}}{X6_{t-1} + X12_{t-1}} - X14_{t-2} + 10}{X6_{t-1} + X12_{t-1}}$
AG1	$\frac{-0.54 * X13_{t-1} * X13_{t-2} + (-0.50 * X9_{t-4} + 0.48) * X10_{t-1} * X14_{t-2}}{X10_{t-1} * X14_{t-2}}$
AG2	$-0.36 * X9_{t-4} + \frac{X9_{t-4} + X10_{t-3}}{X9_{t-3} + 5.55 * X9_{t-4}} + 2.01$
AG3	$-0.50 * X9_{t-4} - \frac{X9_{t-4}}{X11_{t-1}} + \frac{X9_{t-1} + X9_{t-4} + X10_t}{X10_{t-1} * (-0.50 * X9_{t-4} + X15_t)}$
AG4	$\frac{(-0.50 * X9_{t-4} + 1.37) * (-0.50 * X9_{t-4} + X11_{t-4}) + 3.60}{-0.50 * X9_{t-4} + X11_{t-4}}$
MAL	$-0.05 * X9_{t-4} + \frac{X10_{t-3}}{\frac{X9_{t-3} * X10_{t-3} * X13_t}{X11_{t-2} X14_{t-2}}} + X14_{t-1} - 0.50$
FEM	$-0.50 * X9_{t-4} - 0.50 * \frac{X9_{t-4} * X12_{t-1}}{X15_{t-4}} - \frac{2.01}{0.25 * X9_{t-4}}$

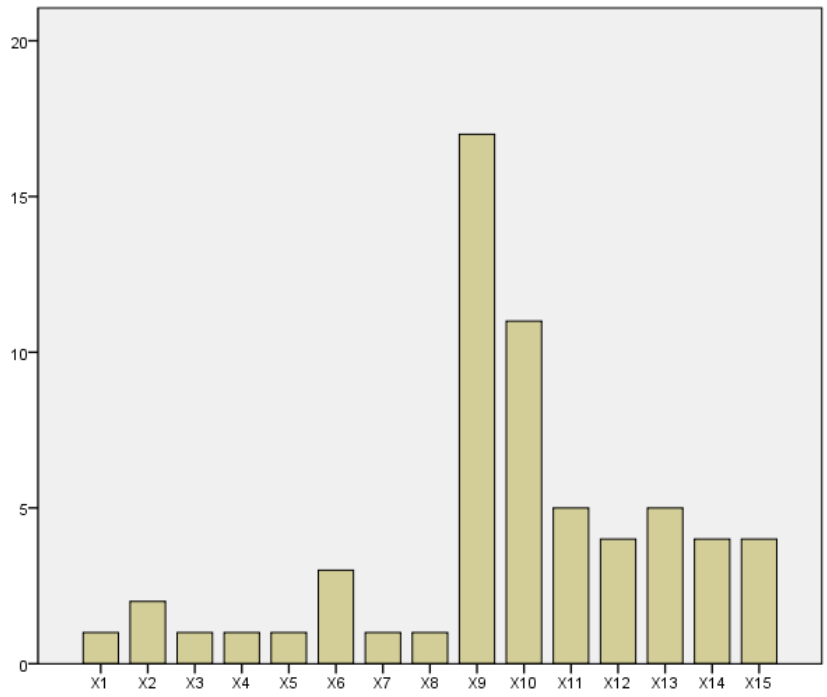
Notes: See Table 2 for the codification of each group of consumers.

Table 4 (cont.)
CCIU for each socio-demographic group

ED1	$-0.36 * X9_{t-4} + 0.05 * \frac{X10_{t-2}}{\frac{X19_{t-4} + X15_{t-4}}{X12_t}} + 2.01$
ED2	$\frac{X9_{t-3} + (-0.36 * X9_{t-4} + 2.01) * (X7_{t-2} - 4.20 * X10_t - X10_{t-4} - 5.01)}{X7_{t-2} - 4.20 * X10_t - X10_{t-4} - 5.01}$
ED3	$-0.50 * X9_{t-4} + \frac{X10_{t-3}}{1.46 + X13_{t-4} - X9_{t-4} * \left(2.01 + \frac{2.92}{X9_{t-4}}\right)} + 2.01$
PR8	$\frac{X2_{t-4}}{X6_{t-1} + X9_{t-4}} - 0.50 * X9_{t-4} + \frac{X11_t}{X10_t}$
PR9	$-0.41 * X9_{t-4} - \frac{5.16}{X9_{t-4} * \frac{X12_t - 2.01}{X6_{t-3}}}$
PR0	$(X1_{t-1} + 2.01 * X9_{t-4}) * \frac{X9_{t-4}}{-9.01 * X9_{t-4} + X10_t + 2.77}$

Notes: See Table 2 for the codification of each group of consumers.

Fig. 2. Histogram with mean frequency for each survey variable



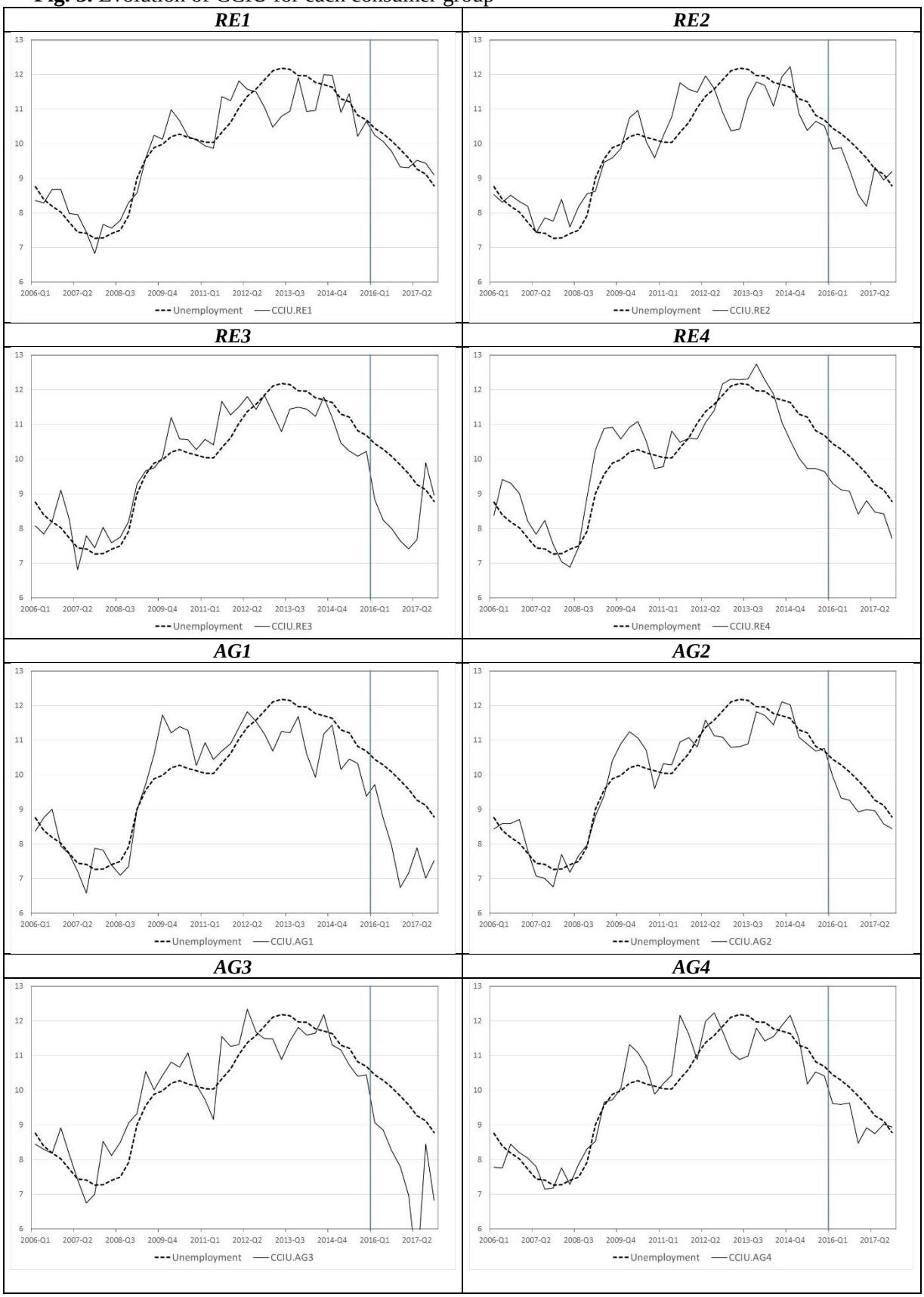
Next, we use the last two years of the sample period corresponding to 2016 and 2017 to assess the out-of-sample forecasting performance of the CCIU for each group of consumers. Due to data availability, PR0, PR8, PR9 are just evaluated for the first two quarters of 2016. In Table 5 we present the MAFE and the RMSFE for each socio-demographic group. The lowest forecast errors are obtained for the households within the first quartile regarding income (RE1), the respondents with secondary education (ED2) and the unemployed (PR0). These results are in line with those obtained by Sorić et al. (2019), who found that unemployed respondents' expectations were most accurate that those obtained by employed respondents. Behavioural economics offers a plausible explanation for this pattern through the "availability heuristic" (Kahneman and Tversky, 1974). It seems that consumer groups with the highest odds for negative outcomes on the job market perceive, update, and process economic information much more intensively, which in our case results in more accurate unemployment predictions. The only exception to this pattern is income segregation, where the highest income group is more precise.

Table 5
Out-of-sample forecast accuracy (2016.Q1–2017.Q4) – Socio-demographic groups

	MAFE	RMSFE		MAFE	RMSFE
RE1	0.30	0.31	ED1	0.58	0.71
RE2	0.64	0.79	ED2	0.30	0.47
RE3	1.58	1.72	ED3	0.89	1.00
RE4	1.01	1.03	MAL	1.25	1.38
AG1	1.83	1.96	FEM	1.00	1.04
AG2	0.62	0.66	PR0	0.10	0.12
AG3	2.05	2.31	PR8	0.93	0.94
AG4	0.59	0.70	PR9	1.41	1.41

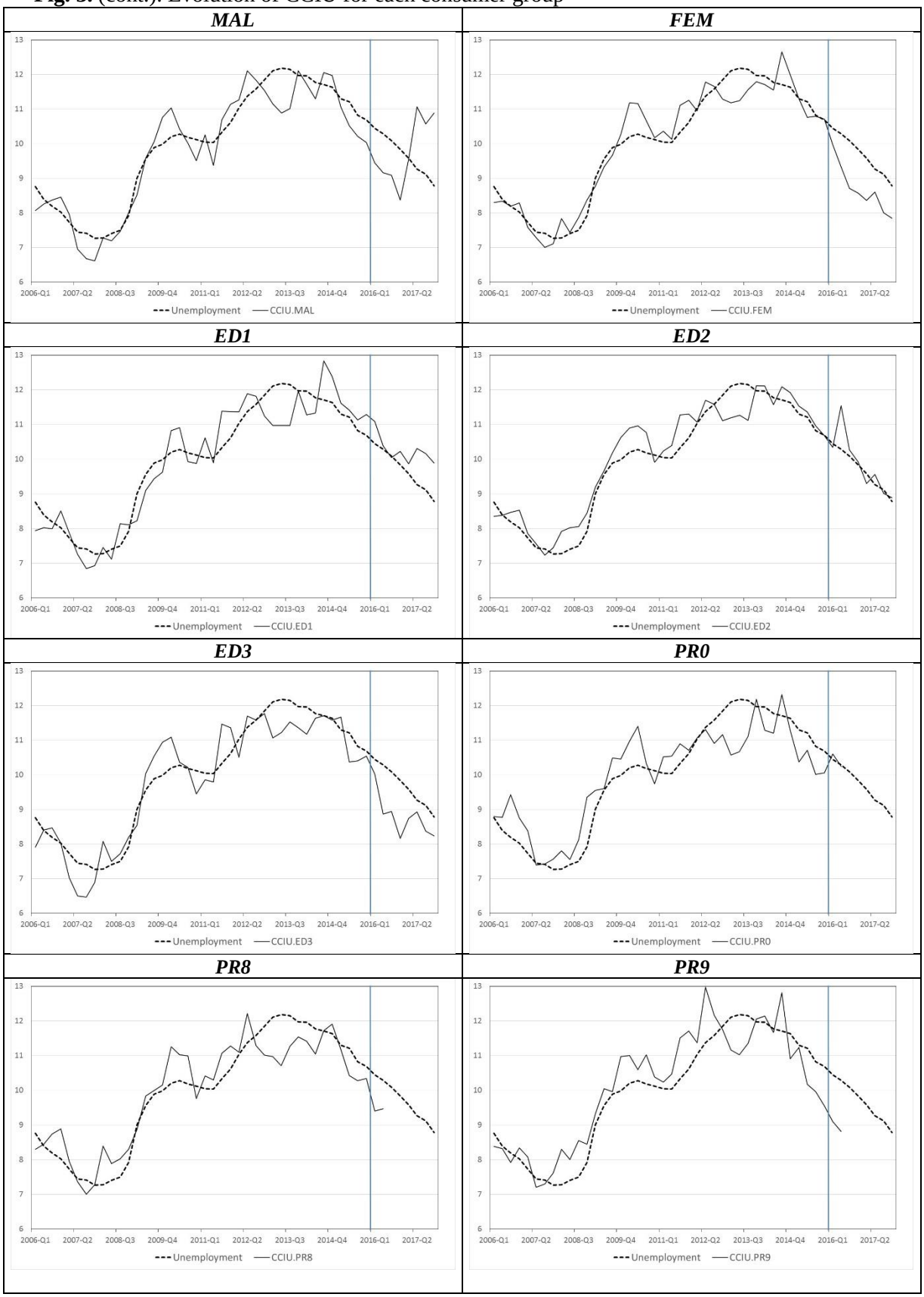
In Fig. 3 we graphically compare the evolution of the CCIU for each consumer group with that of the unemployment rate in the EA. We can observe how the out-of-sample performance of the evolved indicators ostensibly varies across the different groups of consumers.

Fig. 3. Evolution of CCIU for each consumer group



Notes: See Table 2 for the codification of each group of consumers. The black line represents the evolution of the unemployment confidence indicator (CCIU) for each socio-demographic group of consumers and the dashed black line the evolution of the unemployment rate. The vertical line in 2016:Q1 marks the beginning of the out-of-sample period.

Fig. 3. (cont.). Evolution of CCIU for each consumer group



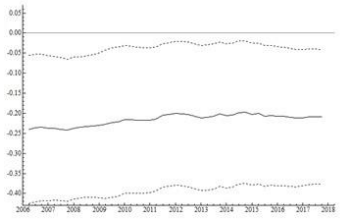
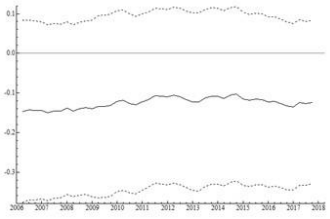
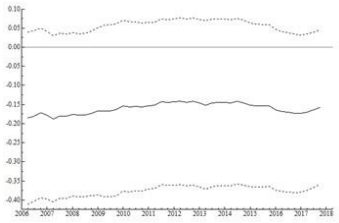
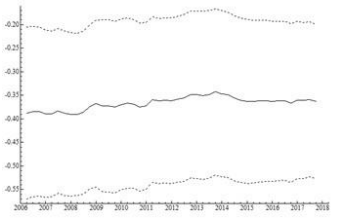
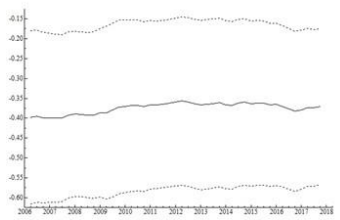
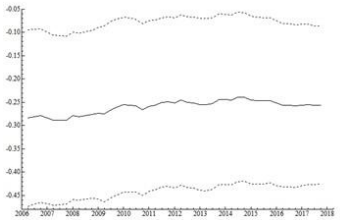
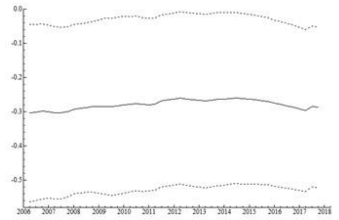
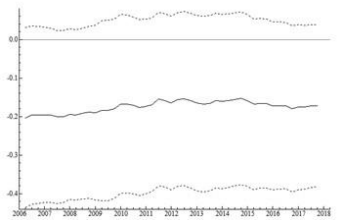
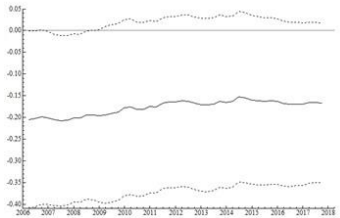
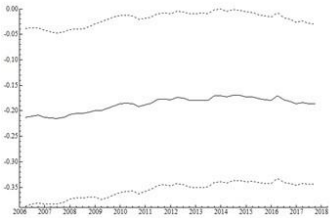
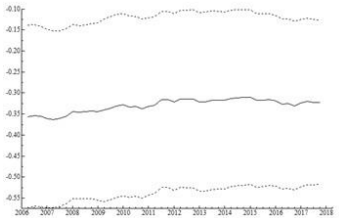
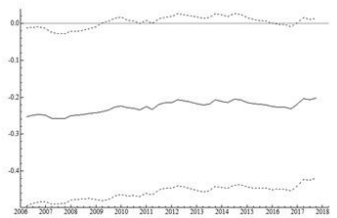
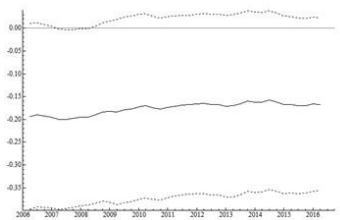
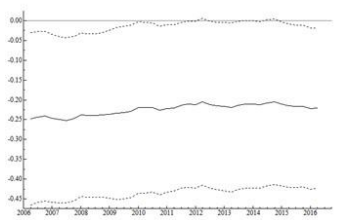
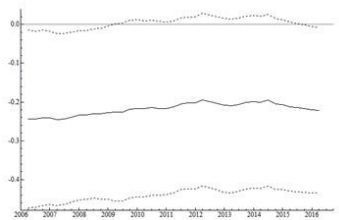
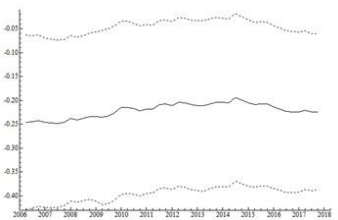
Notes: See Table 2 for the codification of each group of consumers. The black line represents the evolution of the unemployment confidence indicator (CCIU) for each socio-demographic group of consumers and the dashed black line the evolution of the unemployment rate. The vertical line in 2016:Q1 marks the beginning of the out-of-sample period.

Finally, we estimate expressions (2) and (3) for the 2006:Q2–2017:Q4 period for all five CCIU potential determinants (GDP, private consumption, unemployment, HICP, and stock prices) and for all socio-demographic groups. Due to data availability, PR8, PR9, and PR0 models are estimated for the 2006:Q2- 2016:Q2 period. As the influence of private consumption and HICP were not found to be significant, the evolution of their corresponding time-varying parameters has been omitted. The full set of results, together with their corresponding diagnostic test results, are available upon request. We hereby only present the results for the models comprising GDP, unemployment and stock prices in Fig. 4, 5 and 6 respectively.

It should be pointed out that the concept of a confidence interval is regarded as a mirror image of a significance test. A 95% confidence interval not including zero implies that the parameter at hand is significant at the 5% significance level. We find that the analysed GDP parameter, despite being time-varying, does not reveal any considerable structural breaks or substantial changes in the nature of the underlying relationship. On the contrary, the effect of GDP on CCIU is quite stable throughout the business cycle. The only exception is observed for the three occupational categories. The full-time employed consumers (PR8) respond to changes in the state of the economy almost in the entire analysed period. On the other hand, part-time workers (PR9) and the unemployed (PR0) are able to detect signals only in times of very low economic uncertainty, such as the period before the 2008 crisis.

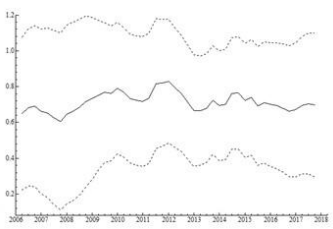
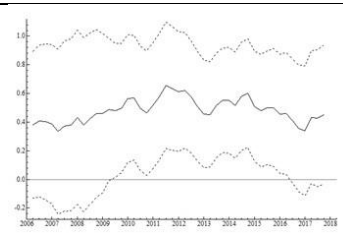
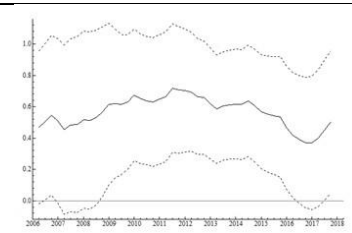
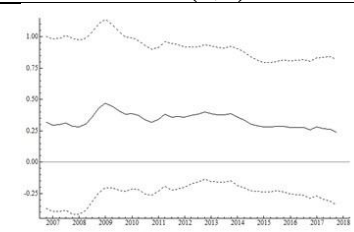
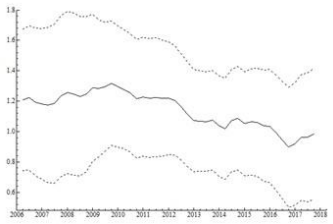
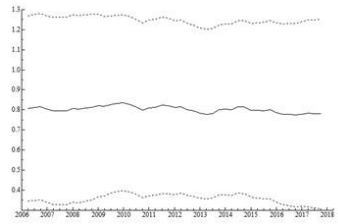
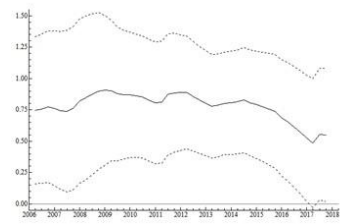
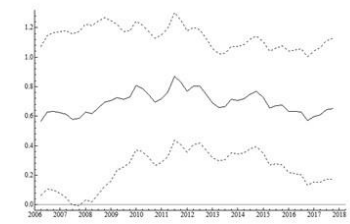
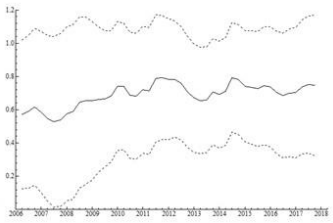
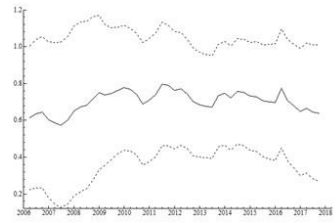
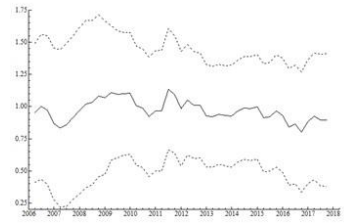
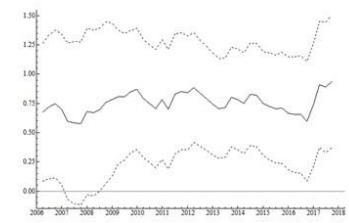
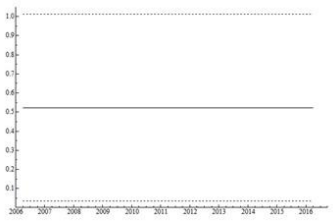
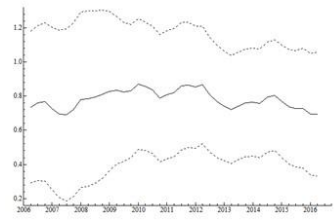
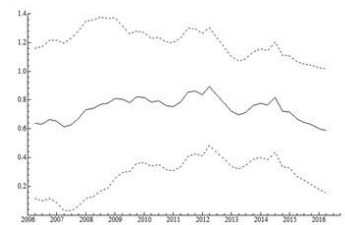
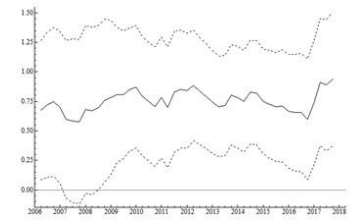
However, out of the five assessed macroeconomic determinants of CCIU, GDP is the one that shows a more heterogeneous behaviour among the 16 observed socio-demographic groups. When it comes to stratification according to income, GDP resonates only in the CCIU of the wealthiest and the lowest income groups. The tails of the income distribution seem to be highly responsive to the overall state of the economy. The central part of the income distribution does not corroborate such findings. Regarding the three examined educational levels, only the least educated (ED1) do not respond to GDP. Age does not seem to alter the way GDP feeds into consumers' unemployment sentiment. All four age groups reveal a significant, negative, and rather stable relationship between the two observed variables. Finally, women seem to be more responsive to GDP than men.

Fig. 4. Time varying effects of GDP on CCIU by socio-demographic group

RE1	***RE2***	***RE3***	***RE4***
			
AG1	***AG2***	***AG3*** (N,H)	***AG4***
			
ED1 (N)	***ED2***	***ED3***	***MAL***
			
PR0	***PR8***	***PR9***	***FEM***
			

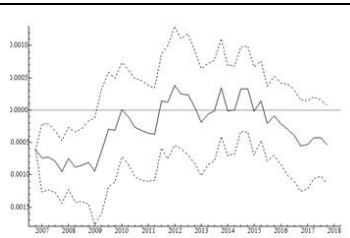
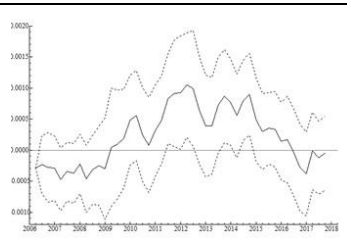
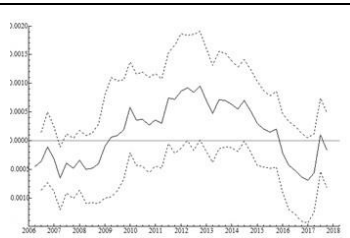
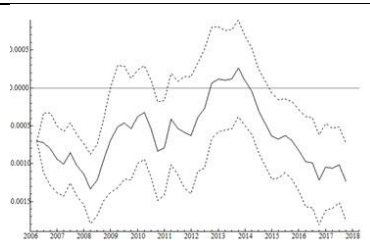
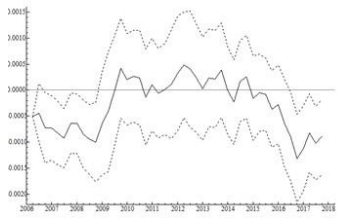
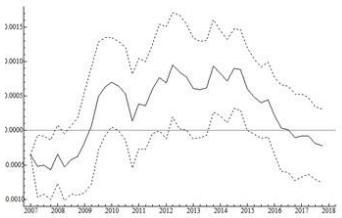
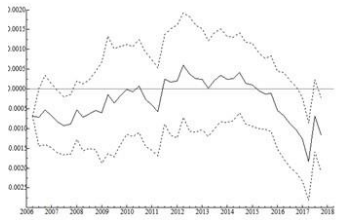
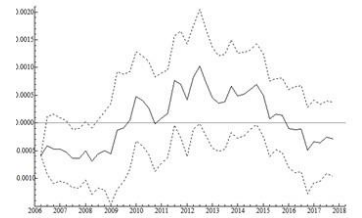
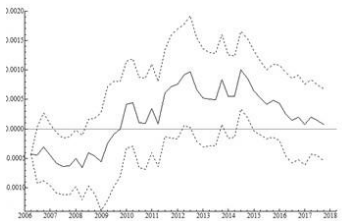
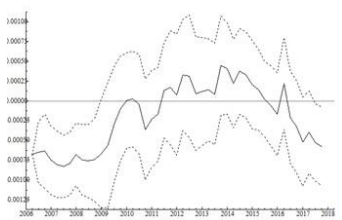
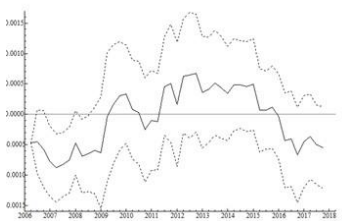
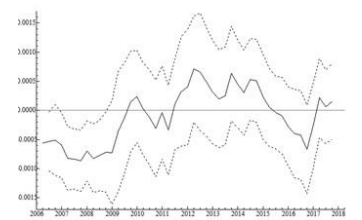
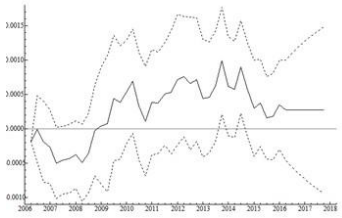
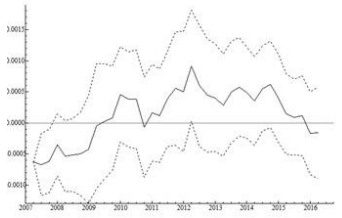
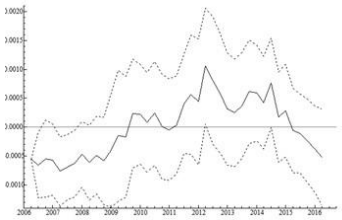
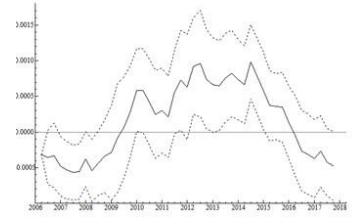
Notes: N, H, and A respectively imply non-normality, heteroskedasticity, and 4th order autocorrelation at the 5% significance level.

Fig. 5. Time varying effects of unemployment on CCIU by socio-demographic group

RE1	***RE2***	***RE3***	***RE4*** (H,A)
			
AG1	***AG2***	***AG3*** (N)	***AG4***
			
ED1 (N)	***ED2***	***ED3***	***MAL*** (N)
			
PR0	***PR8***	***PR9***	***FEM***
			

Notes: N, H, and A respectively imply non-normality, heteroskedasticity, and 4th order autocorrelation at the 5% significance level.

Fig. 6. Time varying effects of stock prices on CCIU by socio-demographic group

RE1	***RE2***	***RE3***	***RE4***
			
AG1	***AG2***	***AG3*** (N,H)	***AG4***
			
ED1	***ED2*** (H)	***ED3***	***MAL***
			
PR0	***PR8***	***PR9***	***FEM***
			

Notes: N, H, and A respectively imply non-normality, heteroskedasticity, and 4th order autocorrelation at the 5% significance level.

The influence of unemployment on the CCIU is found to be statistically significant, of positive sign, and highly constant throughout the observed period. The only exception is the lowest income category of consumers (RE4), where no significant effect of actual unemployment is found. There is no particular evidence that the global financial crisis and the subsequent sovereign debt crisis have had a substantial effect on the way actual unemployment affects the CCIU. This finding is striking taking into account the documented unemployment hysteresis in the EA (Krištić et al., 2018). The most interesting finding is that the magnitude of unemployment effects is independent of consumers' socio-economic status, as it is quite similar for all 16 subgroups.

A general tendency observed from Figure 6 is that stock market results feed into consumers' unemployment sentiment only in a stable economic environment. As expected at the theoretical level, a significant negative relationship is found before the 2008 crisis. During the crisis and the subsequent period of high political and economic uncertainty in the EA, the assessed parameter fades away.

The findings of our state space analysis are particularly relevant from the perspective of a specific branch of literature dealing with the determinants of consumer sentiment. Namely, Garner (1981) was among the very first economists to focus not on the leading characteristics of consumer confidence, but on the driving forces of confidence itself. Garner (1981) identified a set of macroeconomic variables that explained a large part of the variability in US consumer confidence. In that sense, the author interpreted consumer confidence as a mere reflection of the macroeconomy, deprived of any particular added value in forecasting economic activity.

However, more recent research by Beltran and Durre (2003) and Lahiri and Zhao (2016) finds evidence on the contrary. Our results add to this empirical debate by establishing socio-demographics as the moderating variable in the link between consumer expectations and unemployment. We show that the CCIU closely tracks the macroeconomy on the aggregate level, but closer examination of different subgroups of consumers offers a more compelling story. In Garner's (1981) sense, the CCIU offers a unique set of information non-related to HICP and private consumption. This finding is extremely robust and confirmed for all socio-demographic strata. As far as GDP is concerned, we observe quite heterogeneous results across the socio-demographic groups. Again, it seems that for the oldest subgroup, the least educated stratum, and the unemployed, the CCIU offers a unique transmission mechanism and does not solely reflect the evolution of the economy.

The impact of stock market returns on the CCIU is not that diverse with regards to socio-demographics, but it is highly time-varying. The general tendency is that the CCIU picks up on stock market information in times before the 2008 global financial crisis. However, the stated relationship breaks after that. This finding adds to the branch of the literature conjecturing that the role of expectations and other similar psychological concepts, such as uncertainty, considerably increase before recessions. One of the initial contributions to this literature was given by Garner (1991), who postulated that consumer expectations significantly improved macroeconomic predictions prior to an abrupt and unanticipated recession. Recent evidence of the usefulness of agents' sentiment in macroeconomic has been provided by Baker et al. (2016) and Christiansen et al. (2014) *inter alia*. In that sense, the 2008 global crisis and the subsequent unemployment hysteresis in the euro area cannot be fully explained using financial and macroeconomic data.

4. Conclusion

In this study we have used a novel approach to design a quarterly consumer confidence indicator of unemployment in the euro area for sixteen socio-demographic groups. With that aim we have exclusively used consumer expectations as an input. By means of genetic programming we have evolved a wide range of survey expectations to obtain the combination of expectations that best tracks the evolution of unemployment. We have analysed the out-of-sample forecasting performance of the evolved expression and found that the proposed indicators outperform both the scaled balance and the autoregressive forecasts for most groups of consumers.

We have found that the survey questions regarding expected major purchases over the next year, mostly lagged four quarters, and savings at present, both contemporaneously and lagged, are by far, the variables that most frequently appeared in the evolved expressions of the different consumer groups, hinting at their predictive potential to track the evolution of unemployment.

When analysing the differences in forecast accuracy across groups, we have obtained better predictions for respondents comprised in the first quartile with regards to the income of the household and respondents with at least secondary education. The group

where we observed the bigger differences among categories is the occupation, where the lowest forecast errors are obtained for the unemployed respondents.

When analysing the determinants of unemployment sentiment across consumer groups, we have found that for women, the oldest, the unemployed and the groups with lower education, consumer confidence unemployment indicators follow a trajectory quite independent of the fundamentals. This finding suggests that future research on the determinants of consumer confidence would benefit from an in-depth socio-demographic stratification, rather than an aggregate level analysis.

Finally, it has to be noted that the proposed approach to generate consumer confidence unemployment indicators is a data-driven method and therefore lacks any theoretical background. Additionally, there have been several issues left for further research. First, the implementation of alternative evolutionary algorithms to test if they can improve the forecast accuracy of empirically-generated estimates of unemployment expectations. Second, the extension of the analysis to the different European countries in order to examine the similarities between the obtained functional forms across countries. A final issue left for further research is a thorough identification of the political and economic circumstances that enhance and deteriorate consumers' unemployment views.

References

Álvarez-Díaz, M. (2019). Is it possible to accurately forecast the evolution of Brent crude oil prices? An answer based on parametric and nonparametric forecasting methods. *Empirical Economics*. Forthcoming.

Azariadis, C. (1981). Self-fulfilling prophecies. *Journal of Economic Theory*, 25(3): 380–396.

Baker, S. R., Bloom, N. and Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636.

Ball, L., Jalles, J. T. and Loungani, P. (2015). Do forecasters believe in Okun's Law? An assessment of unemployment and output forecasts. *International Journal of Forecasting*, 31: 176–184.

Barnichon, R., and Nekarda, C. J. (2012). The ins and outs of forecasting unemployment: Using labor force flows to forecast the labor market. *Brookings Papers on Economic Activity*, Fall 2012: 83–131.

Beltran, H. and Durre, A. (2003). The determinants of consumer confidence: the case of United States and Belgium, *Center for Operations Research and Econometrics Discussion Papers* 2003053, Université Catholique de Louvain.

Bluedorn, J., Aiyar, S., Duval, R., Furceri, D., Garcia-Macia, D., Ji, Y., Malacrino, D., Qu, H., Siminitz, J. and Zdzienicka, A. (2019). Strengthening the Euro Area: The role of national structural reforms in enhancing resilience. *IMF Staff Discussion Note*, SDN/19/05.

Chen, S. H. and Kuo, T. W. (2002). Evolutionary computation in economics and finance: A bibliography. In S. H. Chen (Ed.), *Evolutionary Computation in Economics and Finance* (pp. 419–455). Heidelberg: Physica-Verlag.

Claveria, O. (2019a). A new consensus-based unemployment indicator. *Applied Economics Letters*, 26(10): 812–817.

Claveria, O. (2019b). Forecasting the unemployment rate using the degree of agreement in consumer unemployment expectations. *Journal for Labour Market Research*, 53(3): 1–10.

Claveria, O., Monte, E. and Torra, S. (2017). 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. and Torra, S. (2018). A data-driven approach to construct survey-based indicators by means of evolutionary algorithms. *Social Indicators Research*, 135(1): 1–14.

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

Couch, K. A. and Fairlie, R. (2010). Last hired, first fired? Black-white unemployment and the business cycle. *Demography*, 47(1): 227–247.

Christiansen, C., Eriksen, J. N. and Moller, S. V. (2014). Forecasting US recessions: The role of sentiment. *Journal of Banking & Finance*, 49: 459–468.

Dabhi, V. K. and Chaudhary, S. (2015). Empirical modeling using genetic programming: a survey of issues and approaches. *Natural Computing*, 14(2): 303–330.

Doornik, J. A., and Hansen, H. (2008). An omnibus test of univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, 70(s1), 927–939.

Eurostat (2019). *News release*, 30/08/2019. Available at: https://ec.europa.eu/eurostat/news/news-releases

Garner, C. A. (1981). Determinants of consumer sentiment. *Journal of Business Research*, 9(2): 205–220.

Garner, C. A. (1991). Forecasting consumer spending: should economists pay attention to consumer confidence surveys? *Federal Reserve Bank of Kansas City Economic Review* May/June: 57–71.

Hutter, C. and Weber, E. (2015). Constructing a new leading indicator for unemployment from a survey among German employment agencies. *Applied Economics*, 47(33): 3540–3558.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2): 263–292.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1), 35–45.

Koopman, S. J., Harvey, A. C., Doornik, J. A., and Shephard, N. (1999). *Structural time series analysis, modelling, and prediction using STAMP*. London: Timberlake Consultants Press.

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

Koza, J. R. (1995). Genetic Programming for econometric modeling. In S. Goonatilaje and P. Treleaven (Eds.), *Intelligent Systems for Finance and Business* (pp. 251–269). London: John Wiley and Sons.

Krištić, I. R., Dumančić, L. R. and Arčabić, V. (2018). Persistence and stochastic convergence of euro area unemployment rates. *Economic Modelling*, 76: 192–198.

Lahiri, K. and Zhao, Y. (2016). Determinants of consumer sentiment over business cycles: Evidence from the US surveys of consumers. *Journal of Business Cycle Research*, 12(2): 187–215.

Parker, R. E. and Rothman, P. (1997). The current depth-of-recession and unemployment-rate forecasts. *Studies in Nonlinear Dynamics & Econometrics*, 2(4): 151–158.

Perić, B. and Sorić, P. (2018). A note on the Economic Policy Uncertainty Index. *Social Indicators Research*, 137(2): 505–526.

Rossi, B. and Sekhposyan, T. (2017). Macroeconomic uncertainty indices for the Euro Area and its individual member countries. *Empirical Economics*, 53(1): 41–62.

Smith, R. E., Vanski, J. E. and Holt, C. C. (1974). Recession and the employment of demographic groups. *Brookings Papers on Economic Activity*, 1974(3): 737–760.

Sorić, P. (2018). Consumer confidence as a GDP determinant in new EU member states: A view from a time-varying perspective. *Empirica*, 45(2): 261–282.

Sorić, P., Lolić, I., Claveria, O., Monte, E., and Torra, S. (2019): Unemployment expectations: A socio-demographic analysis of the effect of news. *Labour Economics*, 60, 64–74.

White, D. R., McDermott, J., Castelli, M., Manzoni, L., Goldman, B. W., Kronberger, G., Jaśkowski, W., O'Reilly, U. and Luke, S. (2013). Better GP benchmarks: community survey results and proposals. *Genetic Programming and Evolvable Machines*, 14(1): 3–29.

Xu, H. and Couch, K. A. (2017). The business cycle, labor market transitions by age, and the great recession. *Applied Economics*, 49(52): 5370–5396.

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