

A new metric of consensus for Likert-type scale questionnaires: An application to consumer expectations

Oscar Claveria*

AQR-IREA, University of Barcelona (UB)

Abstract. In this study, we present a metric of consensus for Likert-type scales. The statistic provides the level of agreement for any given number of response options as the percentage of consensus among respondents. With this aim, we use a geometric framework that allows us to analytically derive a positional indicator. The statistic is obtained as the relative weight of the distance from the point containing the proportions of observations that fall in each category to the centre of a regular polygon with as many vertices as categories, which corresponds to the point of maximum dissent. The polygon can be regarded as the area that encompasses all possible answering combinations. In order to assess the performance of the proposed metric of consensus, we conduct an iterated forecasting experiment to test whether the inclusion of the degree of agreement in households' expectations improves out-of-sample forecast accuracy of the unemployment rate in seven European countries and the Euro Area. We find evidence that the level of consensus among households contains useful information to predict unemployment rates in all cases. This result shows the potential of agreement metrics to track the evolution of economic variables. Finally, we design a simulation experiment in which we compare the sampling distribution of the proposed metric for three- and five-response alternatives, finding that the distribution of the former shows a higher level of granularity and dispersion.

Keywords: Likert scales; agreement; geometry; consumer surveys; expectations; unemployment

JEL Classification Codes: C14; C51; C52; C53; D12; E24

Acknowledgements

I would like to thank the Editor and two anonymous referees for their useful comments and suggestions. Earlier versions of this paper were presented at the 5th *International Conference on Time Series and Forecasting* (ITISE 2018, Granada, September 19-21) and the 39th *International Symposium on Forecasting* (ISF 2019, Thessaloniki, June 16-19). We want to thank the participants of the sessions for their helpful comments and suggestions.

* Corresponding author. Tel.: +34-934021825. Department of Econometrics, Statistics and Applied Economics, University of Barcelona, Diagonal 690, 08034 Barcelona, Spain. *E-mail address:* oclaveria@ub.edu

1 Introduction

The widespread practice of online shopping has fostered the use of customer satisfaction surveys to gather feedback from consumers. Most of these questionnaires make use of Likert scales to elicit the degree of satisfaction of web users. Likert scales were developed to measure people's attitudes [1], and they are the most common approach to scaling responses in survey research. Likert scales result when survey participants are asked to rank their agreement with a set of items on a scale that has a limited number of possible responses that are presented in a sequence. Although the number of responses is usually five, which can take the following form: "strongly agree", "agree", "undecided", "disagree", "strongly disagree", it may vary. See [2] for a discussion on the optimal number of response alternatives.

As a result, Likert scales are ubiquitous in opinion polls. Consumer surveys, which are conducted among households, use Likert-type questionnaires with three and five reply options. These surveys are the main source for eliciting the expectations of consumers. Consequently, household survey-based expectations are widely used as explanatory variables in quantitative forecasting models [3–4] and also to test economic hypothesis [5–8].

The aim of the paper is twofold. On the one hand, we present a metric to compute the degree of consensus among respondents in Likert-type questionnaires for any given number of responses. The measurement of agreement (and dissent) is a key component of data analysis and the study of human behaviour [9]. Although most metrics of qualitative variation are designed so that the extreme categories have a greater weight, in this study we propose a metric that equates all categories, making the neutral or no-change values have the same importance in the calculation. By means of a geometric framework, we derive a measure of positional agreement that has a straightforward interpretation, as it provides the percentage of agreement among the respondents. In addition, the positional nature of the indicator allows monitoring the evolution of the response pattern in a regular polygon with as many vertices as answering categories, which contains all possible response combinations.

On the other hand, we assess the performance of the proposed consensus statistic. With this aim, we first compute the level of agreement between households regarding their expectations about the future evolution of unemployment. The analysis is performed

for a set of long-term member states of the European Union (EU) that reflects the diversity in terms of the evolution of unemployment in European countries, especially since the 2008 financial crisis. Then, we evaluate whether the metric helps to improve the accuracy of unemployment rate forecasts.

With this objective, we design an iterated one-period ahead forecasting experiment in which we generate pseudo out-of-sample predictions of the unemployment rates using an autoregressive model as a benchmark. Next, we replicate the experiment including the agreement metric as a predictor, and test if the reduction in the mean percentage absolute forecast error is statistically significant.

The structure of the paper is as follows. The next section introduces consumer surveys. We then present the methodological approach. Empirical results are provided next. Finally, we end with some concluding remarks.

2 Survey data on households' expectations

Households' expectations are not directly observable, and therefore are elicited through surveys. In the United States, the University of Michigan Survey Research Center started conducting Surveys of Consumers in 1946, covering three broad areas of consumer sentiment: personal finances, business conditions in the economy as a whole, and the appraisal of market conditions for large household durables, vehicles and houses. Based on the significance of consumers' subjective assessments of the economic and social trends, the European Commission conducts the harmonised Consumer Survey since May 1972. The joint harmonised EU Consumer Survey is comprised of twelve monthly questions and three quarterly questions. Its main purpose is to collect information on households' spending and savings intentions, and to assess their perception of the factors influencing these decisions. To this end, consumers are asked about the household financial situation, the general economic situation, savings, and intentions with regard to major purchases [10].

The Consumer Survey mainly consists of qualitative questions, following a similar answer scheme in which responses are given according to a three-option ordinal scale ("increase", "remain unchanged", "decrease") or a five-option ordinal scale ("sharp increase", "slight increase", "constant", "slight decrease", "sharp decrease"). Due to their simplicity, the questionnaires are easy to complete. Surveys are conducted during the first

three weeks of each month, and results are published at the end of each month. Consequently, households' expectations are available prior to the publication of official data, making them particularly useful for monitoring economic developments and short-term forecasting.

Since the 2008 Great Recession, there has been a renewed interest in the measurement of economic uncertainty. In [11], the authors found that during times of high uncertainty survey respondents tended to give more heterogeneous answers to the questions focused on relevant economic variables. Since then, economic uncertainty has been increasingly approximated by the degree of disagreement among survey respondents [12–14]. In this sense, in [15], the authors proposed using the cross-sectional variation of individual expectations. With the objective of capturing this heterogeneity, different measures have been considered, including the index of qualitative variation (see [16–17]).

In this study, instead of measuring the level of disagreement, the focus is on the level of agreement among survey respondents. Another difference with previous research lies in the fact that while most measures of disagreement are exclusively built by means of the responses that fall into the extreme categories (“increase” and “decrease”), while we incorporate the information coming from the respondents expecting a variable to remain constant. With the aim of assessing the proposed metric of consensus, we use households' expectations about the future evolution of unemployment, elicited from the joint harmonised EU Consumer Survey. In the questionnaire, households are asked how they expect the number of people unemployed in the country to change over the next twelve months. In the Appendix, we present the results of a simulation experiment in which we compare the sampling distribution of the proposed metric with that of the index of qualitative variation for different numbers of response options.

3 Methodology

In this section we present a methodology to compute a metric of agreement among survey respondents of Likert-type questionnaires. The frame of reference is based on a geometric application to determine the likelihood of disagreement among election outcomes proposed by [18]. Let us assume a Likert-type questionnaire with K reply options, where $R_{i,t}$ denotes the aggregate percentage of responses in category i at time t , where $i = 1, \dots, K$ and $t = 1, \dots, n$.

Given that the sum of the shares of reply options adds to 100, the vector X_t containing all the information from the surveyed units ($R_{i,t}$), which corresponds to the barycentric coordinates, can be represented as a point on a regular polygon [19]. This polygon, within which all possible combinations of response options are contained, has as many vertices as answering categories. Each vertex corresponds to a point of maximum consensus. We propose measuring the level of agreement as the ratio between ‘the distance of the point to the barycentre’ and ‘the distance from the barycentre to the nearest vertex’. See Figure 1 in the next section for two specific cases ($K = 3$ and $K = 5$).

Since the barycentric coordinate system allows computing the vertical distance of a point in the polygon to the nearest edge, a positional measure of consensus at a given time period t can be formalised as:

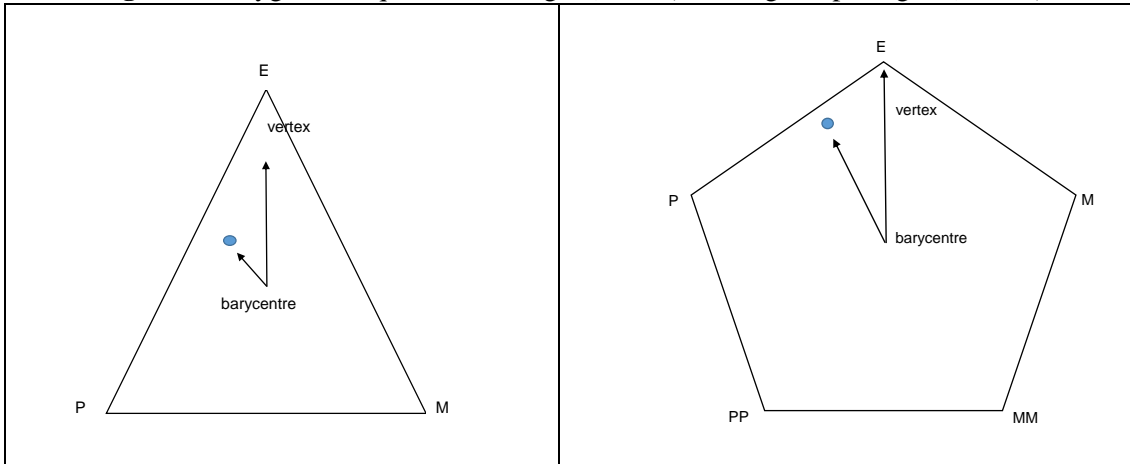
$$C_t = \sqrt{\sum_{i=1}^K \left(R_{i,t} - \frac{100}{N} \right)^2} / \sqrt{(K-1)/K} \quad (1)$$

This metric reaches the maximum of 100% when a response category draws all the responses (total consensus), and the minimum value of zero when the answers are evenly distributed across the K response categories (maximum dissent).

4 Empirical results

The empirical analysis is done for households’ survey data regarding their unemployment expectations. We use monthly data from the Consumer Survey conducted by the European Commission (https://ec.europa.eu/info/business-economy-euro/indicators-statistics/economic-databases/business-and-consumer-surveys_en). The sample period goes from January 2006 to December 2017. Consumers are asked whether they expect a certain variable to “sharply increase”, “slightly increase”, “remain constant”, “slightly decrease” or “sharply decrease”. We respectively denote the aggregated percentages of the individual replies in each category as PP_t , P_t , E_t , M_t and MM_t . Consumers are also faced with questions with three reply options, in which they are asked whether they expect a variable to “increase”, “remain constant” or “decrease”. In this case, the percentages of respondents are respectively noted as P_t , E_t , and M_t . These shares configure vector X_t , which is represented with a grey point in the polygons in Figure 1, where we depict the resulting polygons for both three- and five-response options.

Figure 1. Polygons – Equilateral triangle ($K = 3$) vs. Regular pentagon ($K = 5$)



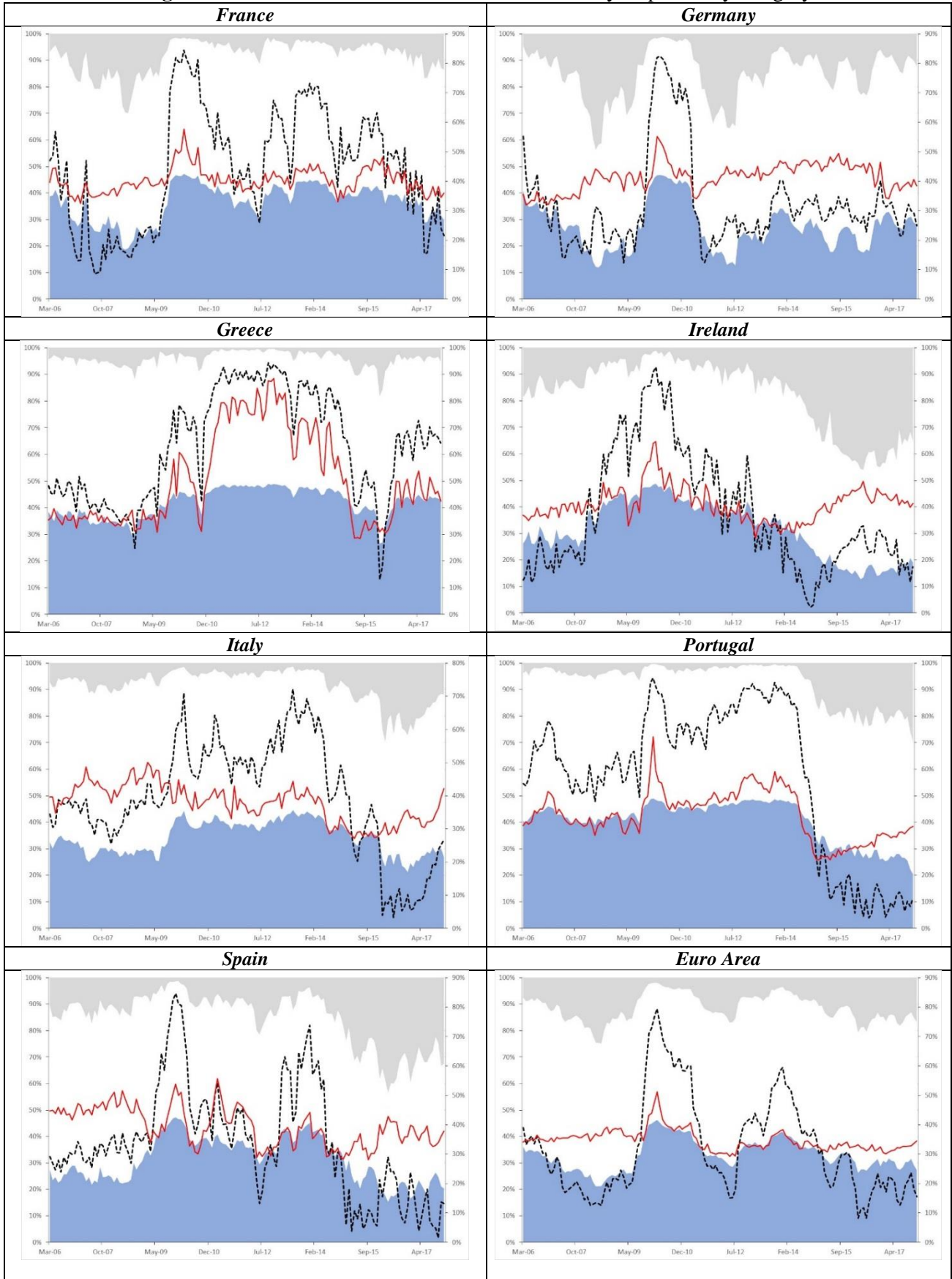
Note: The equilateral triangle corresponds to the three-reply option, where E denotes the % of “remains constant” replies, P the % of “increase”, and M the % of “decrease”. The regular pentagon corresponds to the five-reply question, where E denotes the % of “remains constant” replies, P the % of “slight increase”, PP the % of “sharp increase”, MM the % of “sharp fall”, and M the % of “slight fall”. The grey point in the polygons corresponds to a unique convex combination of all reply options for a given period in time.

In this study, we compare the performance of the proposed metric for both scenarios: the percentage of agreement for three ($C3_t$) and for five reply options ($C5_t$). As unemployment expectations are elicited via five reply options, in order to compute $C3_t$, we opt for grouping all positive responses in P_t , all negative ones in M_t , and incorporating the “do not know” share in E_t .

In Figure 2, we graph the evolution of both consensus measures for the set of countries analysed in the study (France, Germany, Greece, Ireland, Italy, Portugal, Spain and the Euro Area). As a backdrop we represent the distribution of survey responses grouped in three categories. We can see that both measures co-evolve during the sample period, being the five-response consensus metric the one that shows less dispersion. This notion is further confirmed in Tables A1 and Figure A1 of the Appendix, where we present the results of the simulated sampling distributions of the polygons defined in Figure 1 by generating a uniform set of 10,000 vectors.

Several authors have recently addressed the effect of online job searches on unemployment forecasts [20–21], but the role of unemployment expectations has been largely overlooked. To fill this gap, we test if the proposed measure of consensus in unemployment expectations helps to improve the accuracy of unemployment forecasts. With this aim, we use seasonally adjusted unemployment rates provided by the OECD (<https://stats.oecd.org/index.aspx?queryid=36324>).

Figure 2. Consensus metrics and distribution of survey responses by category



Note: Note: The blue area represents the evolution of the percentage of “fall” responses (sharply and slightly) regarding the level of unemployment over the next 12 months, while the light grey area represents the % of “increase” responses (sharply and slightly) and the white area the % of “remain the same (no-change)” responses. The black dashed line represents the evolution of the of the consensus measure for three categories, while the red line the consensus for five categories.

In order to evaluate the forecasting performance of both metrics, we introduce the percentage of agreement among households regarding the future evolution of unemployment. We use both $C3_t$ and $C5_t$ as explanatory variables in autoregressive (AR) models. This model is usually referred to as ARX or dynamic regression model. AR models explain the behaviour of the endogenous variable as a linear combination of its own past values:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (2)$$

Where y_t refers to the rate of unemployment at period t , and ε_t to the innovation, which is assumed to behave as a white noise process. Regarding the number of lags that should be included for each period in every country, we choose between models with a minimum of 1 lag and a maximum of 4 lags, selecting the model with the lowest Akaike's information criterion [22].

We design an iterated one-period ahead forecasting experiment in order to assess the forecast accuracy of the out-of-sample predictions. We use the last twelve periods to compute the mean absolute percentage forecast error (MAPFE), which is a scale-independent measure that weighs the absolute forecast error by the actual value of the variable for every point in time:

$$MAPFE = \frac{100}{n} \sum_{t=1}^n \left| \frac{e_t}{y_t} \right| \quad (3)$$

Where e_t denotes the forecast error at period t . The fact that we are dealing with positive data and comparing countries with different unemployment rates, makes the MAPFE particularly suitable in this case (see [23]). We also compute the small-sample modification of the Diebold-Mariano (DM) test of forecast accuracy proposed by [24] to evaluate whether the inclusion of the consensus metric significantly lowers forecast errors. 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.

Finally, in order to provide further insight regarding the forecasting performance of both $C3_t$ and $C5_t$, we also compute the % of periods with lower absolute error (PLAE). This accuracy measure was proposed by [25], and is a dimensionless measure based on the CJ statistic for testing market efficiency [26]. It can be regarded as a generalisation of the 'percent better' measure proposed by [27] to compare the forecast accuracy of the models to a random walk. The statistic consists on the proportion of periods in which the model under evaluation obtains a lower absolute forecast error than the benchmark model.

Given two competing models A and B , where A refers to the forecasting model under evaluation, whereas B stands for benchmark model, the PLAE can be computed as:

$$PLAE = \frac{\sum_{t=1}^n \lambda_t}{n} \text{ where } \lambda_t = \begin{cases} 1 & \text{if } |e_{t,A}| < |e_{t,B}| \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

In this study we use the AR model as a benchmark. Table 2 contains the results of the pseudo out-of-sample forecasting evaluation.

Table 2. Out-of-sample forecast accuracy – MAPFE and DM test values (2017.01-2017.12)

	MAPFE			% PLAE	
	AR	ARX with $C3_t$	ARX with $C5_t$	ARX with $C3_t$	ARX with $C5_t$
France	1.273	1.046 (0.732)	1.109 (1.245)	75.0%	75.0%
Germany	0.734	0.718 (0.209)	1.014 (-1.748)	50.0%	41.7%
Greece	1.222	0.905 (3.179)	0.773 (4.827)	91.7%	100%
Ireland	2.404	0.516 (4.283)	0.949 (2.111)	91.7%	66.7%
Italy	1.102	0.924 (1.912)	1.400 (-1.925)	75.0%	25.0%
Portugal	2.431	1.062 (11.039)	2.599 (-0.941)	100%	33.3%
Spain	1.579	0.814 (5.683)	1.513 (1.065)	100%	66.7%
Euro Area	1.121	0.374 (14.188)	1.160 (-1.136)	100%	33.3%

Note: MAPFE stands for the mean percentage absolute forecast error. Modified DM test statistic with NW estimator in parentheses. Null hypothesis: the difference between the two competing series (AR vs. ARX) is non-significant. A negative sign of the statistic implies that the second model has bigger forecast errors. The PLAE measures the number of out-of-sample periods with lower absolute errors than the benchmark model (AR model).

We can observe that the information coming from the degree of agreement helps to refine unemployment rate predictions in all countries. This improvement is statistically significant in most of the countries for the model augmented with the metric of agreement for $K = 3$. The reduction in forecast errors seems more relevant in countries with high unemployment rates (Greece, Ireland, Portugal, and Spain). These results are in line with those obtained by [28–31], who found that households' expectations exhibit a high forecasting accuracy. To compare the forecasting performance between $C3_t$ and $C5_t$, we use the % of PLAE and find that, with the exception of Greece, in the rest of the countries the three-response metric performs better.

5 Concluding remarks

The paper presented a new measure of consensus for Likert-type questionnaires. We proposed using a geometric framework to construct a positional indicator that gives the degree of agreement as the relative weight of a distance. This ratio ponders the distance of each point, containing the proportions of observations that fall in each category, to the centre of a regular polygon, which has as many vertices as response options. As a result, the proposed metric provides the percentage of agreement for each item, regardless of the number of categories. Hence, the study contributes to the literature through the development of an innovative approach to measure consensus in Likert scale questionnaires, which provides researchers and practitioners with an additional tool to analyse the information coming from web-customer satisfaction surveys and household qualitative surveys.

With the aim of assessing the performance of the proposed metric of agreement, we used it to compute the level of consensus in consumers' unemployment expectations in seven European countries and the Euro Area. We designed a two-step iterated forecasting experiment in which we first generated out-of-sample predictions of the unemployment rates using an autoregressive model as a benchmark. Then we replicated the experiment augmenting the models with the agreement metric as a predictor so as to test whether its inclusion significantly improved forecast accuracy.

We found that the degree of agreement improved forecast accuracy in all cases. The reduction in forecast errors was found to be statistically significant in four of the countries included in the analysis (Greece, Ireland, Italy and Spain) and in the Euro Area. On the one hand, this finding reveals that the degree of consensus in households' unemployment expectations contains useful information to predict unemployment rates. On the other hand, these results underline the importance of the measurement of agreement, and hint at the usefulness of consensus-based metrics to track the evolution of economic variables.

References

1. Likert R (1932) A technique for the measurement of attitudes. *Archives of Psychology* 140: 1–55. <https://psycnet.apa.org/record/1933-01885-001>
2. Cox EP (1980) The optimal number of response alternatives for a scale: A review. *Journal of Marketing Research* 17(4): 407–422. <https://doi.org/10.2307/3150495>
3. Barnes ML, Olivei GP (2016) Are survey expectations theory-consistent? The role of central bank communication and news. *Journal of Money, Credit and Banking* 49(5): 1031–1058. <https://doi.org/10.1016/j.euroecorev.2016.01.010>
4. Berge TJ (2018) Understanding survey-based inflation expectations. *International Journal of Forecasting* 34(4): 788–801. <https://doi.org/10.1016/j.ijforecast.2018.07.003>
5. Binder CC (2015) Whose expectations augment the Phillips curve?. *Economics Letters* 136: 35–38. <https://doi.org/10.1016/j.econlet.2015.08.013>
6. Binder CC (2017) Measuring uncertainty based on rounding: New method and application to inflation expectations. *Journal of Monetary Economics* 90: 1–12. <https://doi.org/10.1016/j.jmoneco.2017.06.001>
7. Dräger L, Lamla MJ, Pfajfar D (2016) Are survey expectations theory-consistent? The role of central bank communication and news. *European Economic Review* 85: 84–111. <https://doi.org/10.1016/j.euroecorev.2016.01.010>
8. Rosenblatt-Wisch R, Scheufele R (2015) Quantification and characteristics of household inflation expectations in Switzerland. *Applied Economics* 47(26): 2699–2716. <https://doi.org/10.1080/00036846.2015.1008773>
9. Von der Gracht HA (2012) Consensus measurement in Delphi studies. Review and implications for future quality assurance. *Technical Forecasting & Social Change* 79(8): 1525–1536. <https://doi.org/10.1016/j.techfore.2012.04.013>
10. European Commission (2020) *A user guide to the Joint Harmonized EU Programme of Business and Consumer Surveys*. Brussels: Directorate-General for Economic and Financial Affairs. https://ec.europa.eu/info/files/user-guide-joint-harmonised-eu-programme-business-and-consumer-surveys_en
11. Bachmann R, Elstner S, Sims ER (2013) Uncertainty and economic activity: Evidence from business survey data. *American Economic Journal: Macroeconomics* 5(2): 217–249. <https://doi.org/10.1257/mac.5.2.217>
12. Claveria O (2020) Uncertainty indicators based on expectations of business and consumer surveys. *Empirica*. Forthcoming. <https://doi.org/10.1007/s10663-020-09479-1>
13. Dibiasi A, Iselin D (2019) Measuring Knightian uncertainty. *KOF Working Papers* No. 456. ETH Zurich, KOF Swiss Economic Institute, Zurich. <https://doi.org/10.3929/ethz-b-000341984>
14. Girardi A, Reuter A (2017) New uncertainty measures for the euro area using survey data. *Oxford Economic Papers* 69(1): 278–300. <https://doi.org/10.1093/oep/gpw058>
15. Lamla MJ, Maag T (2012) The role of media for inflation forecast disagreement of households and professional forecasters. *Journal of Money, Credit and Banking* 44(7): 1325–1350. <https://doi.org/10.1111/j.1538-4616.2012.00534.x>
16. Mokinski F, Sheng X, Yang J (2015) Measuring disagreement in qualitative expectations. *Journal of Forecasting* 34(5): 405–426. <https://doi.org/10.1002/for.2340>
17. Maag T (2009) On the accuracy of the probability method for quantifying beliefs about inflation. *KOF Working Papers*, No. 230. ETH Zurich, KOF Swiss Economic Institute, Zurich. <https://doi.org/10.3929/ethz-a-005859391>
18. Saari DG (2008) Complexity and the geometry of voting. *Mathematical and Computer Modelling* 48(9–10): 551–573. <https://doi.org/10.1016/j.mcm.2008.05.033>
19. Coxeter HSM (1969) *Introduction to Geometry* (2nd Edition). London: John Wiley & Sons. <https://www.wiley.com/en-es/Introduction+to+Geometry,+2nd+Edition-p-9780471504580>

20. D'Amuri F, Marcucci J (2017) The predictive power of Google searches in forecasting US unemployment. *International Journal of Forecasting* 33(4): 801–816. <https://doi.org/10.1016/j.ijforecast.2017.03.004>
21. Vicente MR, López-Menéndez AJ, Pérez R (2015) Forecasting unemployment with internet search data: Does it help to improve predictions when job destruction is skyrocketing? *Technological Forecasting and Social Change* 92: 132–139. <https://doi.org/10.1016/j.techfore.2014.12.005>
22. Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6): 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
23. Hyndman RJ, Koehler AB (2006) Another look at measures of forecast accuracy. *International Journal of Forecasting* 22(4): 679–688. <https://doi.org/10.1016/j.ijforecast.2006.03.001>
24. Harvey D, Leybourne S, Newbold P (1997) Testing the equality of prediction mean squared errors. *International Journal of Forecasting* 13(2): 281–291. [https://doi.org/10.1016/S0169-2070\(96\)00719-4](https://doi.org/10.1016/S0169-2070(96)00719-4)
25. Claveria O, Monte E, Torra S (2015) A new forecasting approach for the hospitality industry. *International Journal of Contemporary Hospitality Management* 27(7): 1520–1538. <https://doi.org/10.1108/IJCHM-06-2014-0286>
26. Cowles A, Jones H (1937) Some a posteriori probabilities in stock market action. *Econometrica* 5(3): pp. 280–294. <https://doi.org/10.2307/1905515>
27. Makridakis S, Hibon M (2000) The M3-competition: results, conclusions and implications. *International Journal of Forecasting* 16(4): 451–476. [https://doi.org/10.1016/S0169-2070\(00\)00057-1](https://doi.org/10.1016/S0169-2070(00)00057-1)
28. Claveria O (2019a) A new consensus-based unemployment indicator. *Applied Economics Letters* 26(10): 812–817. <https://doi.org/10.1080/13504851.2018.1497846>
29. Claveria O (2019b) A new consensus-based unemployment indicator. *Journal for Labour Market Research* 53(3): 1–10. <https://doi.org/10.1186/s12651-019-0253-4>
30. Lahiri K, Monokroussos G, Zhao Y (2016) Forecasting consumption: The role of consumer confidence in real time with many predictors. *Journal of Applied Econometrics* 31(7): 1254–1275. <https://doi.org/10.1002/jae.2494>
31. 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. <https://doi.org/10.1007/s10663-016-9360-4>

Appendix

We design a simulation experiment to compare the performance of the proposed metric of consensus (dissent) to the index of qualitative variation (IQV). Starting from the simulation of 100,000 aggregated response combinations for five categories, we group the shares of the extreme reply options in order to compute the consensus metrics for $K=3$ and $K=5$ (C3 and C5), and the inverse of the IQV for both scenarios ($1-IQV_3$ and $1-IQV_5$).

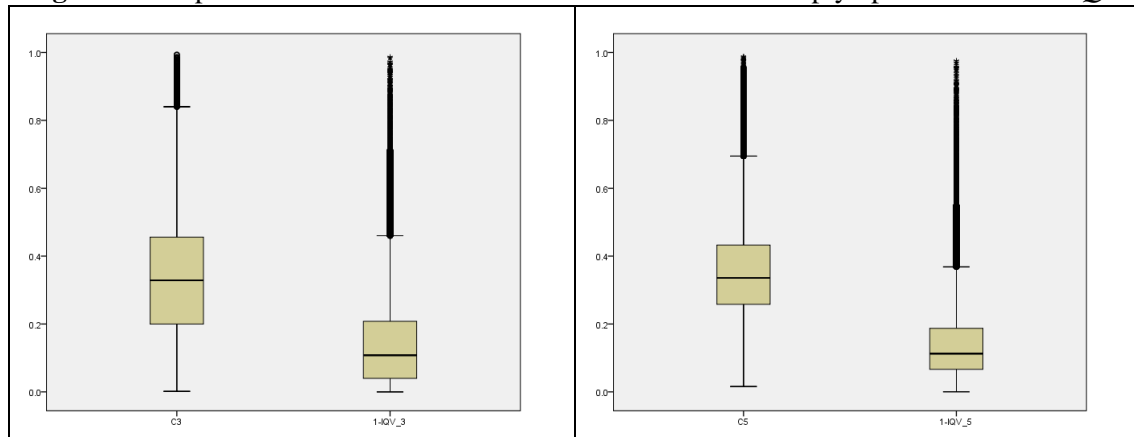
In Table A1 we present the summary statistics for both metrics. While extreme values and standard deviations are quite similar, the interquartile range (IQR), which is obtained as the difference between upper and lower quartiles, is higher for the consensus indicators, especially for three categories. We also observe that the consensus metrics yield higher mean values. This notion is further confirmed in Figure 3, where we compare the boxplots for both measures. The plots represent the respective distributions through their quartiles, without making any assumptions about the underlying statistical distribution.

Table A1. Summary statistics of simulated distribution of C and $1-IQV$

	Mean	Std. Dev.	Min.	Max.	Range	IQR
$C3$	33.797	17.797	0.162	99.337	99.175	25.619
$C5$	35.672	13.377	1.581	98.755	97.174	17.477
$1-IQV_3$	14.590	14.194	0.001	98.677	98.676	16.804
$1-IQV_5$	14.514	11.581	0.025	97.525	97.500	12.070

Note: The range is obtained as the difference between the maximum and the minimum values of the distribution. The IQR refers to the interquartile range, which is obtained by subtracting the first quartile from the third quartile, $Q3-Q1$.

Figure 3. Boxplots of simulated distributions for three- and five-reply options – C vs. $1-IQV$



When comparing the consensus metric for three- and five-response options, it can be seen that the distribution of C3 encompasses a wider range of the scale, and its distribution of scores is more uniform, indicating a higher level of granularity for the median values of the distribution in comparison to C5. Something similar happens when comparing the inverse of IQV for both for $K = 3$ and $K = 5$.

When the proposed consensus measure is compared with the inverse of IQV, we obtain higher average values with the former. Finally, in Figure 4 we use a scatter plot to depict the type of relationship between both metrics, observing a non-linear relationship between them.

Figure 4. Scatter plot of simulated distributions – C vs. $1-IQV$

