Q Open, 2024, **4**, qoae021
https://doi.org/10.1093/qopen/qoae021
Advance access publication date: 9 August 2024
Article

OXFORD

Who can predict farmers' choices in risky gambles?

Henning Schaak [1,*], Jens Rommel [2], Julian Sagebiel [3], Jesus Barreiro-Hurlé [4], Douadia Bougherara [5], Luigi Cembalo [6], Marija Cerjak [7], Tajana Čop [7], Mikołaj Czajkowski [8], María Espinosa-Goded [9], Julia Höhler [10], Carl-Johan Lagerkvist [2], Macario Rodriguez-Entrena [11], Annika Tensi [10], Sophie Thoyer [5], Marina Tomić Maksan [7], Riccardo Vecchio [6] and Katarzyna Zagórska [8]

[1]Department of Economics and Social Sciences, University of Natural Resources and Life Sciences, Vienna, Vienna, Austria
[2]Department of Economics, Swedish University of Agricultural Sciences, Upsalla, Sweden
[3]Biodiversity Economics, German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany
[4]European Commission, Joint Research Centre (JRC), Seville, Spain
[5]CEE-M, Univ. Montpellier, CNRS, INRAE, Institut Agro, Montpellier, France
[6]Department of Agricultural Sciences, University of Naples Federico II, Naples, Italy
[7]University of Zagreb Faculty of Agriculture, Zagreb, Croatia
[8]Faculty of Economic Sciences, University of Warsaw, Warsaw, Poland
[9]Faculty of Economic and Business Science, University of Seville, Seville, Spain
[10]Business Economics Group, Wageningen University & Research, Wageningen, Netherlands
[11]WEARE—Water, Environmental, and Agricultural Resources Economics Research Group, Universidad de Córdoba, Córdoba, Spain

*Corresponding author: Institute of Agricultural and Forestry Economics, University of Natural Resources and Life Sciences, Vienna, Feistmantelstraße 4, 1180 Vienna, Austria. E-mail: henning.schaak@boku.ac.at

Received: January 12, 2024. **Accepted:** July 22, 2024

Abstract

Risk is a pervasive factor in agriculture and a subject of great interest to agricultural economists. However, there is a lack of comprehensive understanding of the knowledge held by farm advisors, students, and economists with regards to farmers' risk preferences. Misconceptions about farmers' willingness to take risks could lead to misguided advice. This study builds upon a recent multinational endeavor that employed a multiple price list to assess risk preferences among European farmers. We expand this research by gathering predictions for farmers' risk preferences from 561 farm advisors, students, and economists. Our objectives are threefold: firstly, we explore variations as to how accurately participants can predict risk preferences in different specializations; secondly, we compare the predictive accuracy of different groups of forecasters; and thirdly, we assess whether modifying incentive mechanisms can improve the accuracy of predictions. Whereas our findings reveal substantial variation in individual predictions, the averages closely align with the observed responses of farmers. Notably, the most accurate predictions were provided by a sample of experimental economics researchers. Furthermore, predictions for different production systems exhibit minimal disparities. Introducing incentive schemes, such as a tournament structure, where the best prediction receives a reward, or a high-accuracy system, where randomly selected participants are compensated for the accuracy of their predictions, does not significantly impact accuracy. Further research and exploration are needed to identify the most reliable sources of advice for farmers.

Keywords: Risk attitudes, Expert predictions, Expert forecasts, Multiple prices lists, Meta-science.

JEL codes: Q12, Q16, C91

1. Introduction

The investigation of farmers' risk preferences is of fundamental importance to agricultural economics (Iyer et al. 2020), as these preferences impact decision-making in various domains such as technology adoption and sustainable farming practices (Dessart, Barreiro-Hurlé and van Bavel 2019). As it is not often possible to estimate risk preferences based on real-world behavior, experimental techniques for determining risk preferences, pioneered by Binswanger (1980), have become widely utilized in Economics in general (e.g. Charness et al. 2013; Friedman et al. 2022) and in Agricultural Economics in particular (e.g. Vollmer et al. 2017; Palm-Forster, Suter and Messer 2019; Bellemare, Lee and Just 2020; Cerroni 2020; Bonjean 2023; Cerroni, Bozzola and Rippo 2023). Despite their inability to predict behavior reliably in all situations (e.g. Hellerstein et al. 2013; Schaak et al. 2017), experimental risk elicitation approaches often do correlate with risk preferences (Menapace et al. 2016; Vollmer et al. 2017). In particular, the use of multiple price lists to elicit risk preferences has faced some criticism (Crosetto and Filippin 2016; Drichoutis and Lusk 2016; Yu et al. 2021) but are still frequently used in agricultural economics (e.g. Iyer et al. 2020; Ruhinduka et al. 2020; Bonjean 2023; Finger, Wüpper and McCallum 2023; Finger et al. 2024) as, among other things, they offer the advantage of enabling the structural estimation of utility functions that can be incorporated into farm-level models (Eisele et al. 2021; Britz 2022; Huber et al. 2022). In conclusion, experimental elicitation methods present the best available method that combines theory with empirics.

Farm advisors and policy-makers need to be able to accurately predict farmers' risk preferences if these are important drivers of farmers' decisions, indeed, if misinterpreted, policies and advice could be misguided. In addition, predictions of research results by various experts can enhance the effectiveness of the research process in social sciences (DellaVigna, Pope and Vivalt 2019). Firstly, predictions provide a systematic approach to elicit prior beliefs from the research community, reducing hindsight bias when research results become available. Predictions can involve practitioners in the experimental design phase (Höhler et al. 2024) thereby establishing a clear benchmark of existing knowledge to challenge prior beliefs and enhance research effectiveness. Secondly, benchmarking predictions ex-ante can facilitate the acceptance of null results, especially when the null deviates from prior beliefs. Thirdly, systematic and regular predictions from an expert community can improve prediction accuracy, informing future research designs and providing valuable insights for policymakers (DellaVigna and Pope 2018b; Milkman et al. 2022; Vivalt and Coville 2023). Fourth, a sample of 'super-forecasters'—experts who are very good at predicting—could be consulted in instances where data collection is not feasible or costly.

While predictions in economics have predominantly focused on laboratory experiments (DellaVigna and Pope 2018a, 2018b), predictions related to experimental outcomes in agricultural economics have been limited to specific topics, such as the behavior of German farmers in a public goods game under different treatments (Rommel et al. 2023b), and use professional academics and graduate students as experts. To our knowledge, there is currently no comprehensive study eliciting knowledge for the crucial topic of risk preferences in European agriculture (Iyer et al. 2020) which would provide insights into system-specific expertise and the potential impact of financial incentives on prediction improvement. Such knowledge would enhance our understanding of more specific expertise beyond social scientists' ability to predict social phenomena (Grossmann et al. 2023) and help farmers identify reliable sources of advice (Wuepper et al. 2021; Rust et al. 2022).

This study has three objectives. Firstly, we aim to explore whether the accuracy of predicting farmers' risk preferences is context-dependent and specifically if expertise differs among farming specializations. Leveraging a recent large-scale replication of an economic experiment with farmers from ten EU member states (Rommel et al. 2023a), we investigate whether risk preferences are easier to predict for wine growers in Croatia, olive farmers in Italy and Spain, potato growers in France, or arable farmers in the Netherlands, Sweden, and Germany. Note that we are bound by the original study which restricts our experimental design and the subsequent results. For instance, while we can estimate differences between the samples, readers should also be aware of confounds. Sampling, payments, and survey modes differed across the original studies (see Rommel et al. 2023a for details). Hence, estimates cannot be causally attributed to farming systems alone, but should be interpreted as a joint estimate of farming system, country, and data collection.

Secondly, we compare different groups of forecasters. Drawing on multiple samples comprising 561 farm advisors, students, and economists (including Polish, French, Croatian, and Italian farm advisors; Swedish agriculture students; a mixed group of experts from Spain recruited from research and policy networks; experimental economists; and a diverse group of experts based on convenience samples), we examine how these groups predict farmers' decisions when facing risky gambles as part of a multiple price list. By deliberately obtaining samples that match expertise and specialization at national levels, we can also investigate whether there are skewed perceptions for the specialization that most closely matches a given forecaster sample's expertise. For example, one would assume Polish farm advisors more knowledgeable about this risk profiles of farmers as compared to farmers in some other group. This also allows us to assess which sample of forecasters could best replace primary data from farmers (if at all).

Thirdly, we seek to enhance our understanding of the optimal approach to incentivize accurate predictions in the future. Previous research on predictions of research results has focused on the impact of reference values, raw units vs. standard deviations, sliders vs. text entry, and different slider bounds on the accuracy of forecasts (DellaVigna, Otis and Vivalt 2020). There is also extensive literature on belief elicitation in Economics (e.g. Trautmann and van de Kuilen 2015; Charness et al. 2021) and Agricultural Economics (e.g. Norris and Kramer 1990; Hardaker and Lien 2010; Cerroni et al. 2023) for which these methodological questions are relevant. Building upon this line of inquiry, we explore another vital question: the role of financial incentives. Financial incentives are commonly used to promote careful decision-making in economic experiments (Camerer and Hogarth 1999; Voslinsky and Azar 2021) which aligns with the objective of eliciting more accurate predictions in our study. We test two tournament scheme incentives against two accuracy-based incentivized systems through a randomized between-subjects design with five conditions and with a control treatment. One out of every 50 participants is selected for payment. In treatments 1 and 2, the selection occurs within a tournament scheme, with low and high rewards, respectively. In treatments 3 and 4, incentives are independent of others' actions and solely based on the deviation from the actual outcome, with low and high penalties for inaccurate predictions, respectively. Treatment 5, the control condition, involves randomly paying a fixed price to one out of every 50 participants, regardless of their predictions.

Table 1. Descriptive Statistics of the participants.

	($N = 561$)
Age	
Mean ± Standard deviation	38.26 ± 11.92
Median	37
Min	20
Max	84
Female	
If respondent is female	240/555 (43.2%)
Professional background	
Economics or Business Studies	184 (32.8%)
Agricultural Sciences/Farming	238 (42.4%)
Other	139 (24.8%)
Sample	
Polish farm advisors	109 (19.4%)
Croatian farm advisors	56 (10.0%)
French farm advisors	72 (12.8%)
Italian farm advisors	51 (9.1%)
Spanish experts	59 (10.5%)
Swedish students	69 (12.3%)
Experimental economists	76 (13.6%)
Other	69 (12.3%)

Source: Own calculations.

2. Data collection, experimental design and approach for data analysis

2.1 Data collection

Data were collected through an online survey conducted from 15th December 2021 to 28th January 2022.[1] The survey was made available in multiple languages, including Croatian, English, French, German, Italian, Polish, and Spanish, and was distributed through various channels such as research networks of the authors, advisor associations, and students. Upon participants' entry into the system, they were welcomed and provided with an introduction to the survey's objectives. The prediction mechanisms were explained and elicited at the survey's outset. Depending on the specific treatment assigned, the incentive mechanism was explained. To assess the salience of their assigned incentive mechanism and participants' understanding of its functioning, participants were later asked to select the one applied in their case from a list of all applied mechanisms. Additionally, participants were asked to provide socioeconomic information. Participants were also asked to assess the perceived difficulty of the prediction task and the confidence in their predictions.

Data were specifically collected from a subset of countries that were part of the large-scale replication study conducted by Rommel et al. (2023a) which was used for our tag-along experiment. We successfully obtained eight samples: farm advisors from Poland, Croatia, France, and Italy; mixed experts from Spain; Swedish agricultural students; experimental economists; and a diverse group labeled as 'Other,' comprising forecasters from various countries with different backgrounds.

All participants provided informed consent and were offered the opportunity to receive a debrief by subscribing to a summary of the research results. Prior to data collection, we pre-registered the basic analysis, and the details can be found at https://aspredicted.org/Z8Z_FV7. A comparison of the pre-registered hypothesis and the analyses presented in this paper is given in Appendix 1. The survey was completed by a total of 561 participants, each of whom predicted the outcomes for all eight samples from the original study.

Consequently, the final dataset encompasses 4,488 predictions. Table 1 presents summary statistics regarding the sociodemographic characteristics of the participants.

2.2 Experimental design

The data to be predicted was obtained from farmers and collected as part of a comprehensive cross-country initiative that aimed to replicate the study by Bocquého, Jacquet, and Reynaud (2014) in various European Union member states (for detailed information, refer to Rommel et al. 2023a). This study, conducted in the latter half of 2021, involved farmers making choices between riskier and safer options using a modified version of the risk preferences elicitation task developed by Tanaka et al. (2010). Notably, farmers in Rommel et al. (2023a) were required to indicate a single switch point from option A to option B.

The data collection for the prediction study occurred after the collection of farmer data but prior to the disclosure of farmers' choice outcomes, taking place between late 2021 and early 2022. Authors of the replication study were excluded from participating in the prediction study. For each of the eight farmer samples (wine growers from Croatia, olive farmers from Italy and Spain, potato growers from France, and arable farmers from the Netherlands, Sweden, and Germany), we requested participants to predict the average row number at which farmers in a specific sample would switch from the safer option A to the riskier option B, using one of the multiple price lists from the risk elicitation task introduced by Tanaka et al. (2010). Participants only predicted one of three lists in order to reduce the complexity of the overall prediction task (meaning that participants only had to carry out 8 instead of 24 predictions, moreover, the focus of the present study is on risk whereas the original task also looked at loss aversion and probability weighting). While it would have been interesting to also elicit forecasts in the loss domain, and to account for probability weighting, we were concerned about the length and complexity of the survey. In addition, this approach presented a unique opportunity to obtain a large multi-country dataset rather than a domain comparison (which would be easier to implement in future research focusing on a single study region/farming system).

In the task used for the forecasts, a higher row number indicates a higher predicted average risk aversion. Farmers who choose the safer option A at least seven times (i.e. that switch after the sixth row) are considered risk-averse.

Predictions were elicited on a scale from 0 (indicating farmers who on average never choose the safer option A) to 12 (indicating farmers who on average always choose the safer option A). This format was deemed the most intuitive by the research team and participants entered their predictions by sliding a marker on a scale with one decimal point accuracy for each of the eight samples of the original study. Table 2 presents the price list used, including the expected payoff difference, which was not revealed to either the forecasters or the farmers participating in the original study.

Our main outcome variable of interest is the accuracy of the predictions. We define the accuracy based on the deviation of predicted value from the one found in the actual sample. As the deviation can be positive and negative, we use the absolute deviation as the accuracy measure. Note that this definition implies that *smaller* values (lower bound at zero) indicate predictions with *higher* accuracy. Recall that we obtained eight predictions per participant (one for each farmer sample). We selected this outcome in order to achieve a simple linear interpretation of the results. In the analysis, the robustness of the main model specification was analyzed an alternative accuracy measure (the squared deviation instead of the absolute). Accurate predictions were incentivized in four out of five treatments, which were implemented between subjects (see Table 3 for an overview including the maximum and minimum possible values for the accuracy-based treatments). In treatment ACCLOW, one randomly selected participant from a group of 50 participants was offered a payment, which was calculated as 300 Euro minus the squared deviation of one randomly selected

Table 2. Multiple price list used in this study and difference in expected value.

Row	Option A		Option B		Expected payoff difference (A–B)
Series 1	Probability 30%	Probability 70%	Probability 10%	Probability 90%	
1	400	100	680	50	77
2	400	100	750	50	70
3	400	100	830	50	60
4	400	100	930	50	52
5	400	100	1,060	50	39
6	400	100	1,250	50	20
7	400	100	1,500	50	–5
8	400	100	1,850	50	–40
9	400	100	2,200	50	–75
10	400	100	3,000	50	–155
11	400	100	4,000	50	–255
12	400	100	6,000	50	–455

Note: Adapted from Tanaka et al. (2010) and Rommel et al. (2023a); displayed units are experimental currency units.

Table 3. Overview of the experimental treatments.

Row	Type	Selection criterion for payment	Maximum and minimum payable amount
ACCLOW	Accuracy	Randomly selected	€300 minus the squared deviation of the prediction from true value (i.e. €300 if deviation is zero; 300 − 144 = €156 if deviation is at the maximum of 12)
ACCHIGH	Accuracy	Randomly selected	€300 minus two times the squared deviation of the prediction from true value (i.e. €300 if deviation is zero; 300 − 288 = €12 if deviation is at the maximum of 12)
TOURHIGH	Tournament	Most accurate prediction	€300
TOURLOW	Tournament	Most accurate prediction	€100
CONTROL	Control	Randomly selected	€300

Source: Own depiction.

prediction out of their eight predictions. In treatment ACCHIGH, the payment was calculated as 300 Euro minus *twice* the squared deviation in order to test for incentive effects, i.e. deviations were punished relatively more in ACCHIGH.

Incentives in the accuracy-based treatments (ACCHIGH and ACCLOW) punished larger deviations from the true value disproportionally through a squared deviation rule. This ensured strictly positive values for the deviations which were subtracted from the €300. The €300 starting value was chosen to ensure positive payments in all instances. Because true values are not close to the extremes in any of the surveys, even the largest deviations resulted in relatively high payments. For example, if the true value was 6, even when choosing an extreme value of 0 or 12 (the lower and upper limits of the scale), then the payment would have been €264 ($300 - 6^2$) in ACCLOW and €228 ($300 - 2 \times 6^2$) in ACCHIGH. Note that by choosing 6 (the center), the maximum possible deviation is also 6 resulting in the same payment even if the true values lie at the extremes of the scale.

In TOURHIGH and TOURLOW (the two tournament schemes), payments of 300 and 100 Euro were offered to the best prediction from a randomly selected sample of a group of

50 participants. In CONTROL, a payment of 300 Euro was offered to a randomly selected participant from a group of 50 participants. Note that 'HIGH' and 'LOW', in the treatment abbreviations, do not only refer to the monetary values for the tournament treatment, but also refer to the level of penalty in the accuracy based one. We received between 100 and 150 responses per treatment. Hence, payments were offered to three participants per treatment for a total of 15 payments.[2]

2.3 Approach for data analysis

To investigate differences across farming systems and groups of forecasters, we used descriptive statistics, visualization, and non-parametric tests. We apply non-parametric multi-comparison Kruskal–Wallis tests to investigate whether the predictions and the accuracy of different samples of forecasters come from the same underlying distributions and pairwise, Wilcoxon rank sum test to assess which forecaster samples differ from each other. That is, we test whether some farmers' behavior is easier or more difficult to predict.

To study the effect of the incentive mechanisms (the third objective), the prediction accuracy was initially assessed using a Kruskal–Wallis test and then used as the dependent variable in regression analysis. Here, two dimensions have to be considered: a financial incentive could (simultaneously) (a) improve the average prediction accuracy and (b) reduce the heterogeneity of the prediction accuracy (i.e. its variance). To simultaneously consider both dimensions, a distributional regression framework was applied which is referred to as Generalized Additive Models for Location, Shape and Scale (GAMLSS). The core idea of GAMLSS, introduced by Rigby and Stasinopoulos (2005), is to not only model the expectation of the dependent variable's distributions, but all the parameters of the assumed distribution, e.g. the mean and the variance of a normally distributed variable. This allows for a straightforward extension of the standard linear regression model. A corresponding GAMLSS with a linear predictor equation for the mean and the variance of the distribution can be written as[3]

$$Y \sim N(\mu, \sigma),$$
$$g_1(\mu) = \eta_1 = X_1\beta_1,$$
$$g_2(\sigma) = \eta_2 = X_2\beta_2.$$

Here, Y represents a vector of observations of the independent variable, which is assumed to be normally distributed, conditional on the sets of dependent variables X_1 and X_2. $g_1(\mu)$, and $g_2(\sigma)$ are the link functions for the corresponding linear predictor equations. This regression model can be estimated using maximum likelihood-techniques. As diagnostic plots indicate that the normality assumption for the residuals does not hold for the estimated models, and in order to account for potential correlations of errors within respondents, clustered standard-errors (at the individual level) are applied. These are calculated using a bootstrap-routine and are used in the results section.

3. Results

The descriptive statistics of the predictions, per predicted sample and for all samples, are presented in Table 4. The table also shows the true means of the farmer samples. Based on the first task of Tanaka et al. (2010), farmers can be characterized as slightly risk-seeking on average, with Polish farmers being the most, and Spanish farmers being the least, risk-averse. Note that this characterization serves illustrative purposes, as it changes when structural estimation across all three lottery tasks is performed as in the original contribution of Rommel et al. (2023a).

Means range from 4.74 in Spain to 6.30 in Poland. The differences in the means of up to 1.64 show that there is a rather large heterogeneity in how farmers respond to the multiple

Table 4. Descriptive statistics of the participants' predictions.

Predicted sample	True Mean	Predictions		
		Predicted mean	Median	SD
Sweden	5.70	5.71	6	2.72
Germany	5.71	6.03	6	2.68
Poland	6.30	6.06	6	2.80
Netherlands	5.80	5.94	6	2.74
Spain	4.74	5.92	6	2.74
Italy	4.96	5.80	5.80	2.82
Croatia	6.05	5.58	5.50	2.76
France	5.28	5.89	6	2.62
Pooled predictions	5.61	5.87	5.94	2.01

Notes: Own calculations, true means based on Rommel et al. (2023a).

price lists (cf. Rommel et al. 2023a). In contrast, the means of the predictions differ by 0.48 at most (from 5.58 for Croatia to 6.06 for Poland). Although the predicted means obtained cover the complete, technically possible, range (from 0 to 12) for all predicted samples, the average predicted mean is close to the true mean in most cases. Pooled across all samples, the difference between the true mean and the predicted mean is 0.26. The smallest difference is found for the Swedish sample (0.01), the largest one for the Spanish sample (1.18). When rounding to integers, which corresponds to values representing choices possible in the MPL, the predicted choice only differs from the observed average choice for the Spanish, French, and Italian sample (by one row). Additional plots of the predictions' distributions of the individual farm systems are presented in Appendix 2.

When the predictions are plotted by sample and forecaster groups (see box plots in Appendix 3) the results show some samples for which medians and means of the predictions by forecaster groups fluctuate around the true mean (e.g. Croatia and Germany), whereas others exhibit a pattern of biased predictions (e.g. Spain and Italy). When testing for differences between forecaster groups' predictions by specialization, Kruskal–Wallis tests only indicate statistically significant differences between forecaster groups for the Swedish ($X^2 = 14.08$; $P = 0.050$) and Croatian samples ($X^2 = 22.45$; $P = 0.002$). Additional pairwise, Wilcoxon rank sum tests to indicate that the null hypothesis can only be rejected for the expert-group-pair of experimental economists and Polish Farm advisors in both samples (using the Bonferroni–Holm correction, at the 5 per cent-level). The predictions can be used to calculate the implied risk aversion coefficients (a measure of the degree of relative risk aversion) predicted by the forecasters and are presented in Appendix 4, showing that average theoretical parameter values are fairly similar for original study participants and forecasters.

Table 5 displays the prediction accuracy, defined as the deviation from the sample average, by sample and forecaster group. The last column (*Pooled*) indicates how much the forecaster samples deviate, on average, from the true means across all eight samples of the original study. The first row (*Pooled Predictions*) displays how much, on average, all pooled predictions deviate from the true mean for each of the eight samples of the original study. In other words, low values in the last column indicate high predictive accuracy of a group of forecasters and low values in the first row indicate that a sample is easier to predict accurately. Note that the sample of researchers provided the most accurate predictions on average, whereas the sample of French farmers was the easiest to predict. The range from low to high accuracy is smaller when considering the diversity of predicted samples (from 2.13 for France to 2.41 for Spain, representing a range of 0.30) than when considering the diversity of forecaster samples (from 1.80 for the researchers to 2.70 for the Polish farm advisors, representing a range of 0.90). Formal testing reveals that the average accuracy of the predictions per forecaster group does not come from the same distribution across

Table 5. Exploring absolute deviations of predictions from the true farmers' means.

		Predicted farmer samples								
Forecaster samples	N	Sweden	Germany	Poland	Netherlands	Spain	Italy	Croatia	France	Pooled
Pooled predictions	561	2.21	2.17	2.26	2.20	**2.41**	2.34	2.25	**2.13**	2.25
Farm Advisors Poland	109	2.59	2.89	2.92	2.60	2.74	2.82	2.66	2.35	**2.70**
Experimental economists	76	1.63	1.56	1.88	1.53	2.18	2.08	1.73	1.81	**1.80**
Farm Advisors Croatia	56	2.12	2.09	2.36	2.69	2.52	2.48	2.63	2.22	2.39
Farm Advisors France	72	1.97	1.91	2.11	1.93	1.99	1.90	1.98	1.95	1.97
Farm Advisors Italy	51	2.82	2.46	2.54	2.81	2.60	2.78	2.45	2.05	2.56
Experts Spain	59	2.14	2.19	2.16	2.17	2.41	2.02	2.14	2.08	2.17
Swedish students	69	2.31	2.06	1.98	2.00	2.32	2.28	2.30	2.20	2.18
Other	69	2.09	1.92	1.87	1.93	2.42	2.21	2.08	2.29	2.10

Source: Own calculations. Note: Bold values for highest and lowest absolute deviation across predicted samples and category of forecasters.

Table 6. Average accuracy by incentive treatments.

Treatment	N	Minimum	Q1	Q2/Median	Q3	Maximum	Mean	SD
ACCLOW	116	0.64	1.43	1.94	2.67	6.33	2.16	1.10
ACCHIGH	118	0.42	1.52	2.23	3.11	6.43	2.41	1.28
TOURLOW	107	0.52	1.43	1.95	2.77	6.43	2.13	1.12
TOURHIGH	112	0.47	1.36	1.95	2.79	5.46	2.15	1.06
CONTROL	108	0.42	1.39	2.10	3.15	6.37	2.37	1.33

Source: Own calculations.

all samples of forecasters (Kruskal–Wallis test; $X^2 = 41.01$; $P < 0.001$), which suggests that at least two samples of predictors in our data follow a different distribution. Pairwise Wilcoxon rank sum tests to indicate statistically significant differences between the average predictions of the experimental economists and farm advisors from Poland, Croatia, and Italy, as well as between the farm advisors from France and Poland in addition to Italy (using the Bonferroni–Holm correction, at the 5 percent-level).

Table 6 summarizes the distribution of all predictions for the incentive treatments. Overall, the mean deviations are similar across treatments. A Kruskal–Wallis test ($X^2 = 4.28$; $P = 0.37$) does not reject the null of equal distributions. Differences in the standard deviations are relatively large and pairwise F-tests indicate incompatibility of the data with the null (e.g. testing the difference of the standard deviation of TOURHIGH from CONTROL yields an F-ratio of 0.63 with $P = 0.017$ for the two-sided test). This indicates that incentives may not necessarily lead to different predictions on average but could help improving the reliability of predictions (see Camerer and Hogarth (1999) for a discussion on the effect of incentives on the variation of experimental outcomes depending on effort).

As seen in the numerical example for payments in ACCLOW and ACCHIGH above, high payments could be achieved by choosing any value, and accuracy had no impact on the probability of being selected, whereas in TOURLOW and TOURHIGH, the respondent must do better than other respondents in order to be selected for a payment. Given this feature of the incentive schemes, one would expect that the tournament schemes induce more effort to think carefully about the forecasts. However, we do not find evidence that this has led to differences in forecasts. In addition, more effort in considering the task may not necessarily lead to higher accuracy.

Table 7. GAMLSS regressions with accuracy as dependent variable.

	Model 1		Model 2	
Predictor	μ	σ	μ	σ
Link function	Linear	Log	Linear	Log
(Intercept)	2.3729***	0.5651***	2.7176***	0.7802***
	(0.1367)	(0.0458)	(0.3251)	(0.1273)
ACCHIGH	0.0498	−0.0177	0.0081	−0.0626
	(0.1746)	(0.0641)	(0.1846)	(0.0712)
ACCLOW	−0.2025	−0.0788	−0.2622	−0.1420*
	(0.1639)	(0.0642)	(0.1625)	(0.0657)
TOURHIGH	−0.2148	−0.0855	−0.2914+	−0.1455*
	(0.1630)	(0.0599)	(0.1662)	(0.0685)
TOURLOW	−0.2273	−0.1017	−0.1911	−0.1257
	(0.1676)	(0.0714)	(0.1858)	(0.0792)
Overestimation			0.1911*	0.1368***
			(0.0878)	(0.0401)
Forecaster: Experimental economists			−0.8337***	−0.3975***
			(0.2102)	(0.1064)
Forecaster: Farm_Advisors_Croatia			−0.2785	−0.0898
			(0.1948)	(0.0736)
Forecaster: Farm_Advisors_France			−0.6427***	−0.2429***
			(0.1754)	(0.0646)
Forecaster: Farm_Advisors_Italy			−0.0215	−0.0191
			(0.2003)	(0.0740)
Forecaster: Experts_Spain			−0.4085*	−0.1440+
			(0.1983)	(0.0863)
Forecaster: Swedish_students			−0.5206*	−0.2523**
			(0.2087)	(0.0836)
Forecaster: Other			−0.4685*	−0.2231**
			(0.2080)	(0.0777)
Female			0.1833+	0.0453
			(0.1013)	(0.0427)
Age			−0.0015	−0.0013
			(0.0050)	(0.0020)
Background agricultural sciences/farming			0.0015	−0.0828
			(0.1347)	(0.0562)
Background other			−0.1363	−0.0681
			(0.1451)	(0.0664)
Num. Obs.	4,488		4,408	
Pseudo-R^2	0.014		0.087	
AIC	17,329.22		16,711.22	
Prediction sample FE	Yes		Yes	

Source: Own calculations. Notes: $+P < 0.1$, $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. Clustered standard errors in parentheses.

We implemented a manipulation check on the incentives treatments by asking respondents to correctly identify the incentive scheme they were assigned to after having made their predictions. As seen in Appendix 5, between 50 per cent and 70 per cent of the respondents could correctly identify their exact treatment. An additional 15 per cent could at least identify the correct basic incentive mechanism (tournament or accuracy).

Table 7 presents the regression results that further investigate the effects of financial incentives on the accuracy of predictions. Distributional regression models are estimated, which include linear predictors for both the mean and the variance of the prediction accuracy. The basic specification (Model 1) only includes an intercept, the binary controls for the predicted sample (omitted for brevity), and four dummy variables for the five treatments (reference category = CONTROL). Model 2 adds the covariates to adjust for the samples of forecasters (setting as reference category the group of Polish farm advisors; i.e. the largest participant subgroup) and socioeconomic characteristics. To account for potential asymmetries in predictions, such as over- or respectively under-estimating farmers' average choices, a binary variable *Overestimation* is included, which takes the value 1 when the prediction underlying the calculated accuracy was above the true mean of the respective sample. All variables are included in both predictor equations. Standard errors are clustered at the individual level to account for correlated predictions within participants. For comparison, estimations for a standard linear model (using ordinary least squares estimation for just the distribution's mean) are given in Appendix 6. The main conclusions hold.

The regressions show only small and statistically insignificant effects of the treatments on the mean of the prediction accuracy which supports the results of the previous subsection. The same holds for the effects on the variance of the prediction accuracy, some effects are statistically significant when the model controls for the predictor sample socioeconomic characteristics (Model 2). We also find that accuracy differs by respondent group, namely some forecasters groups (Experimental Economists, Farm Advisors from France, Swedish students and the miscellaneous group 'other') made more accurate and precise forecasts than the largest participant subgroup ('Polish Advisors'). This supports the findings of the non-parametric tests presented above. Gender, age, and professional background showed no statistically significant effects on prediction accuracy. Finally, the results of the binary variable '*Overestimation*' indicate that predictions were less accurate and noisier in cases where a participant predicted a value larger than the true value for a given farmer group. While it may be surprising that experimental economists perform better than farm advisors in the study, familiarity with multiple price lists could be one reason to explain this result.

In order to assess the robustness of the estimates, alternative specifications were taken into consideration. The general findings were confirmed when estimating the models using alternative specifications for treatment variables (variables indicating the treatment type, accuracy- or tournament-based, and whether an individual was treated or not) or using an alternative distributional assumption for accuracy (Gamma-distribution). The same held when using the squared deviation instead of the absolute deviation as a measure of accuracy. When analyzing only subsets of the individuals that were able to correctly identify their treatments or treatment types (cf. Appendix 5), the same general effect patterns were found, albeit not at statistically significant levels which is probably due to smaller sample size.[4]

It is interesting to understand whether knowledge of the predicted specialization is important to the results. In the absence of detailed data regarding the participants' knowledge of the eight different specializations, we assumed that participants know the specialization of the country they reside in best. This allowed for the creation of a variable indicating whether a prediction for a given farmer sample was made by a forecaster from the corresponding country. Considering the subset of participants assigned to a country-specific forecaster sample, simple regression analyses (see Appendix 7) indicated that predictions were less accurate when the prediction was made for the participants' country. There are no statistically significant interactions between the 'Own country'-dummy variable and the treatment dummies, giving no indication that this bias was mitigated by the financial incentives. We can only speculate as to the reasons for this counter-intuitive result, but it could be that forecasters gave more careful consideration to the sample that they felt most familiar with. If this hampered more intuitive reasoning (for instance applying heuristics and just choose the middle), this might have worsened rather than improved accuracy for this particular study.

Finally, the collected data also allow for some exploratory analyses. In the survey, forecasters were asked to state their confidence in their predictions on a scale from 0 (not confident at all) to 100 (very confident). Interestingly, the forecasters with the most accurate predictions (experimental economists and farm advisors from France) have the lowest average confidence in their predictions (cf. Appendices 8 and 9). A linear fit suggests that, overall, forecasters with higher confidence tend to provide less accurate predictions (recall that the lower values of the measure indicate higher accuracy; see Appendix 9).

4. Discussion

The predictions of farmers' choices exhibit significant heterogeneity, as observed in the raw predictions and the calculated prediction accuracy. However, average predictions align more closely with the actual values reported by Rommel et al. (2023a), indicating a possible 'wisdom of the crowd effect' (DellaVigna and Pope 2018a). Hence, it would be valuable to consider eliciting uncertainty or probability distributions instead of solely focusing on point estimates in the future. This approach can provide a more comprehensive understanding of forecasters' confidence and the range of potential outcomes.

Forecasters tend to make less accurate predictions when forecasting the behavior of farmers from their own country. This discrepancy may be attributed to the forecasters relying on frugal heuristics (Gigerenzer and Brighton 2009) when making predictions for unfamiliar specializations but employing more thoughtful considerations for predictions in familiar settings, resulting in inferior accuracy when compared to relying on heuristics. When the goal is to fill-in for primary data from farmers, e.g. in modeling approaches, it is therefore better to not necessarily rely on national expertise. However, it should also be kept in mind that the median value of forecasts was 6, and in many instances, this forecast has yielded high accuracy. The value 6 is also in the center of the scale, i.e. simply by clicking in the middle, and without giving it too much thought, one could have done relatively well. This would then be an artifact of the results and study design rather than a general advantage of heuristics (unless farmers in the original study have also applied such heuristics).

The findings connect to the stated preferences literature. In the methodological discussion of choice experiments, in bias the elicited value is commonly referred to as validity and efficiency as commonly referred to reliability (Rakotonarivo, Schaafsma and Hockley 2016; Cerroni et al. 2023). The findings regarding average prediction accuracy do not reveal statistically significant effects of monetary incentives on the average accuracy, i.e. the reliability of the predictions. However, when tested against the control, one of the tournament- and one of the accuracy-based incentive schemes lead to statistically significant lower the variances of the forecast accuracy, as indicated by the negative coefficient estimates for the sigma of the treatment dummies. We consider this result to indicate that the validity of the predictions potentially improves through financial incentives. As this result if not found for all treatments, further research is warranted to explore the reliability and validity of predictions more comprehensively and to assess whether the potential effects are practically relevant. Using the above definitions, we cannot conclude that validity and reliability are different across elicitation methods in our sample.

It would also be valuable to investigate whether tournament-based incentives exhibit gender-heterogeneous treatment effects when compared to accuracy-based incentives in terms of accuracy or variation (Niederle and Vesterlund 2007) in order to avoid biased forecasts. Due to the likelihood of the current research being underpowered to detect such effects, future studies should focus on optimizing the utilization of available sample pools to increase statistical power.

The observed differences in accuracy between overestimations and underestimations of true values imply that participants who predicted higher levels of risk aversion made less accurate and more inconsistent predictions. It is important to note that the true mean in most samples falls within the risk-seeking domain for the task, which could contribute to the observed differences in estimation.

Limited sampling possibilities and potential self-selection of respondents are a constraint to the generalizability of our findings. This concern exists at two levels: that of the forecaster groups making the predictions and that of the samples of farmers whose behavior was predicted. Whether the interpretation of the results holds for the underlying population depends on whether the results of Rommel et al. (2023a) are assumed to be representative of the broader population. If this assumption does not hold, the interpretation is limited to the specific samples studied by Rommel et al. (2023a) which poses an additional challenge for forecasters. Additionally, it is important to acknowledge the general critique of multiple price list elicitation methods, including the complexity of the tasks asked to interviewees (Dave et al. 2010; Crosetto and Filippin 2016; Drichoutis and Lusk 2016; Yu et al. 2021), which also applies to the prediction of outcomes.

The information provided to participants regarding the groups of farmers and their specialization was relatively broad, which compelled them to rely on prior knowledge and intuition for their forecasts. Extending this research to investigate the forecasters' ability to predict the outcomes of individual farmers or smaller, more homogeneous groups of farmers would allow us to better understand whether it is task comprehension or the sample that causes low accuracy. Additionally, incorporating quantitative data on past behavior of farmers or farmer groups could enhance our understanding of the differences between intuition-driven and data-driven forecasts (Grossmann et al. 2023). Furthermore, providing a brief summary of the research results to all respondents offers an opportunity to steer interest in the results (Höhler et al. 2024) or to examine whether forecasters update their beliefs after participating in multiple predictions when receiving feedback (Vivalt and Coville 2023).

Student samples as a substitute for farmer samples have been studied in the context of experimental risk elicitation methods (Maart-Noelck and Musshoff 2014; Grüner et al. 2022). Our student subsample, Swedish agricultural science students, represent a group that is easier to sample than advisors. While our primary goal was not to compare the behavior of farmers and students, it has to be noted that the predictions of the student sample do not stand out in comparison to farm advisors or others.

Lastly, it is important to acknowledge that our predictions were based on only one experimental risk preference elicitation task, while the original study employed three multiple price lists to elicit parameters for cumulative prospect theory. By using only one of the lists, we were able to simplify the task for respondents, possibly achieving a larger sample size. However, this approach limits our understanding of other dimensions of risk preferences, such as the degree of loss aversion or probability weighting. Future investigations should explore these aspects, while also considering the potential limitations in the diversity of forecasters that may arise. It is therefore also crucial to delve deeper into how different elicitation formats and the complexity of instructions impact response rates and prediction accuracy.

5. Conclusions

The understanding of outsiders' knowledge regarding farmers' risk preferences has been a neglected area of research. This study aimed to address this gap by analyzing the predictions of 561 agricultural and experimental economics experts, as well as lay people, in a multiple-price-list experiment designed to determine risk preferences for various groups of farmers. This is important in order to understand the differences in expertise regarding a significant behavioral trait of farmers. Should agricultural advisors misconceive farmers' willingness to take risks, they may give misguided advice. The participants in our study were tasked with predicting the average outcome of specific price lists used in the study conducted by

Rommel et al. (2023a) for wine growers in Croatia, olive farmers in Italy and Spain, potato growers in France, and arable farmers in the Netherlands, Sweden, and Germany. The predictions were financially incentivized through five different randomly assigned treatment mechanisms, including two tournament-based schemes, two accuracy-based schemes, and one control treatment with a fixed payment.

By comparing the forecasts with the actual behavioral data of farmers reported by Rommel et al. (2023a), we were able to assess the accuracy of these predictions. We observed variations in the prediction accuracy across different farmer groups, although only a few differences were statistically significant when taking into consideration the diverse expert groups. Interestingly, the predictions were found to be less accurate when participants predicted the behavior of farmers from their own country. We also found that high confidence was associated with poor predictions. This highlights the potential influence of biases associated with participants' prior beliefs when making predictions for their local agricultural context. Our results indicate that the average prediction accuracy is not significantly affected by the different financial incentive mechanisms employed in the study, while incentives appear to have the potential to reduce the variability of predictions, indicating a possible role in improving the consistency and reliability of forecasts.

It is important to acknowledge that this study has certain limitations. Firstly, the generalizability of the findings is subject to the representativeness and external validity of the sampled forecasters and the sampled farmers. Survey-based experiments often face challenges in achieving a comprehensive representation of the underlying population, and potential self-selection biases among respondents may impact the generalizability of the results. The information provided about the farmer groups and their specialization was relatively coarse, requiring forecasters to rely on their prior knowledge and intuition to make predictions. Future research could explore predictions for individual farmers or smaller, homogenous groups, as well as incorporate quantitative data on past behavior to enhance the accuracy of forecasts. In this context, the use of simpler tasks than multiple price lists is warranted, although it might be difficult to find outcomes that are readily and widely available.

In conclusion, this study sheds light on forecasters' knowledge of farmers' risk preferences by analyzing their predictions in a multiple-price-list experiment. The findings highlight variations in prediction accuracy across different farming specializations. Based on the findings of this study, the question arises whom farmers should rely on for risk management advice. While forecasters' predictions showed some accuracy, limitations and biases were evident, particularly when our study participants predicted the behavior of farmers from their own country.

End Notes

1 See the Supplementary information S1.

2 To decide on the winner, groups were divided into equal size (i.e. the actual group size was a bit smaller than 50, which is conceptually equivalent to rounding up the expected value of payments). We successfully contacted and exchanged banking details and executed payments with 10 out these 15 respondents. One respondent explicitly declined the payment, and four others did not respond to our attempt to contact them.

3 Note that GAMLSS is a versatile framework, which allows the incorporation of different effect types (e.g. semiparametric and spatial effects) and complex distributions (with up to four parameters). As these possibilities are not of interest here, we refrain from giving a full introduction to the framework. The interested reader is referred to the canonical references (Rigby and Stasinopoulos 2005; Stasinopoulos and Rigby 2007).

4 All results can be obtained by running code provided in the replication material.

Acknowledgement

The authors thank Sofie Bosma for language editing.

Supplementary material

Supplementary data are available at *Q Open* online.

Funding

M.C. acknowledges financial support from the National Science Centre of Poland (Sonata Bis, 2018/30/E/HS4/00388). J.R. acknowledges financial support from the Swedish Research Council for Sustainable Development, Grant/Award Number: 2019–00993. D.B. and S.T. acknowledge financial support from the University of Montpellier (SRUM) and from the French National Research Agency under the Investments for the Future Program, referred to as ANR-16-CONV-0004 (#DigitAg). R.V. acknowledges financial support through the project PRIN DRASTIC 'Driving The Italian Agri-Food System Into A Circular Economy Model,' PRIN-MIUR—Call 2017, funded by the Italian Ministry of Education, University and Research (MIUR). M.C., T.Č., and M.T.M. acknowledge funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Number 952303 (AgriFoodBoost).

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Data availability

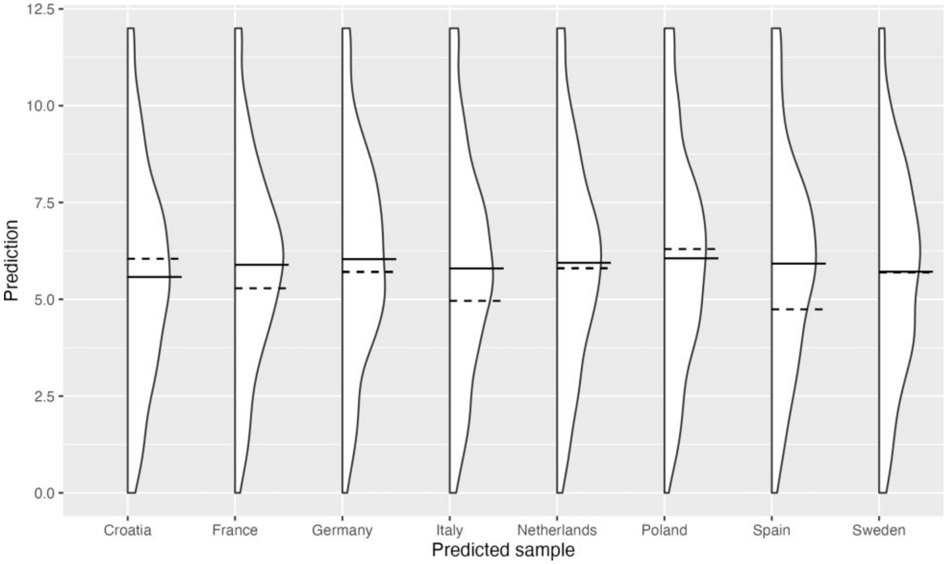
Data and code to replicate all results are available at https://doi.org/10.17605/OSF.IO/D5PT8.

Appendix 1. Comparison between the preregistered and analyses carried out.

Analysis aspect	Preregistration	Paper
Analyses to examine the main question/ hypotheses	–Compare accuracy across conditions using a non-parametric test. –Use a non-parametric test for differences in the distribution of predictive accuracy by the different samples. –Linear panel regression model; dependent variable: accuracy; independent variables: sample dummies, payment treatment dummies (baseline specification); extended specifications including gender, age, education, experience with the sector, experiments, and farmer's behavior, familiarity with any of the samples, and risk and insurance., adjust for the different samples	–Compare accuracy across conditions using a non-parametric test. –Use a non-parametric test for differences in the distribution of predictive accuracy by the different samples. –GAMLSS regression model, in order to also model the of variance of the prediction accuracy; dependent variable: accuracy; independent variables: prediction sample fixed effects, payment treatment dummies (baseline specification); extended specification including gender, age, sample dummies and experience with the sector, clustered standard errors

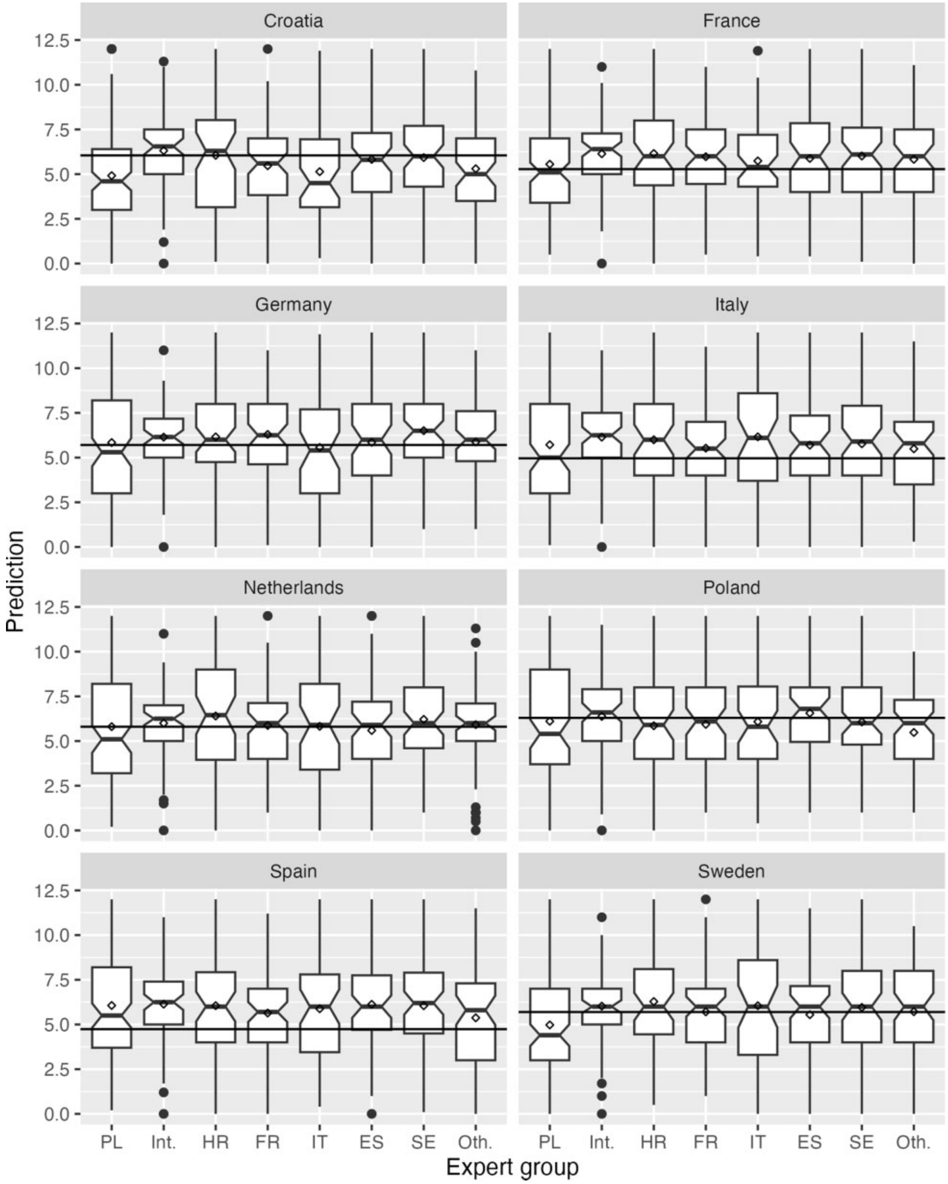
Analysis aspect	Preregistration	Paper
How outliers will be defined and handled	–May remove people who responded very fast or very slow or who showed other irregularities (which may indicate bots). –All analyses for the full sample are reported in an appendix.	–Alternative model specifications: –Linear regression models (Appendix 6) –Different distributional assumption (included the replication code) –Alternative definition of the treatment variables (only treatment type, included the replication code) –Different subsets of the data, based on the ability of the participants to identify their assigned treatment (included the replication code)
Other (secondary) analysis	–None preregistered	–Visualizations of the predictions (Appendices 2 and 3) –Calculated the implied risk aversion coefficients (Appendix 4) –Explorative regressions (Appendix 7) –Explorative analysis of the confidence of the participants in their predictions and the actual predictions (Appendices 8 and 9)

Appendix 2. Distributions of the predictions and observed responses of the different specializations.



Notes: Own calculations, solid lines: means of the predictions, dashed lines: means of the observed responses in Rommel et al. (2023a)

Appendix 3. Predictions by sample and forecaster group.



Notes: Own calculations, additionally to the boxplot-conventions, diamonds indicate the mean, the notches indicate the approximate 95 per cent—interval of the sample-median. Horizontal lines indicate the true means reported by Rommel et al. (2023a).

Appendix 4. *R*-values implied by the predictions.

Country	Average predicted r-value (by experts)	Observed average r-value (by farmers)
Croatia	1.2 $N = 561$	1.12 $N = 104$
France	1.16 $N = 561$	1.06 $N = 96$
Germany	1.14 $N = 561$	1.13 $N = 153$
Italy	1.19 $N = 561$	1.1 $N = 130$
Netherlands	1.16 $N = 561$	1.13 $N = 160$
Poland	1.19 $N = 561$	1.23 $N = 169$
Spain	1.16 $N = 561$	0.99 $N = 130$
Sweden	1.16 $N = 561$	1.05 $N = 218$

Source: Own calculations. Notes: The table shows r values elicited with the midpoint technique (Bocquého, Jacquet and Reynaud 2014) by country. The first column shows the elicited r values from the farmer samples from the original study. The second column shows the r values from the expert predictions based on the predicted switching points. The calculation of the r values are based on Series 1. In Rommel et al. (2023a), the r values are based on all three series which explains the difference between this table and the values reported in Rommel et al. (2023a).

Appendix 5. Control question for treatment mechanism.

Assigned treatment		Answer					
		ACCHIGH	ACCLOW	CONTROL	TOURHIGH	TOURLOW	I don't know
ACCHIGH	N	67	19	8	9	2	13
	%	56.8	16.1	6.8	7.6	1.7	11.0
ACCLOW	N	17	68	9	11	1	10
	%	14.7	58.6	7.8	9.5	0.9	8.6
CONTROL	N	4	12	71	16	0	5
	%	3.7	11.1	65.7	14.8	0.0	4.6
TOURHIGH	N	11	20	8	60	0	13
	%	9.8	17.9	7.1	53.6	0.0	11.6
TOURLOW	N	5	17	10	14	48	13
	%	4.7	15.9	9.3	13.1	44.9	12.1
All	N	104	136	106	110	51	54
	%	18.5	24.2	18.9	19.6	9.1	9.6

Source: Own calculations.

Appendix 6. Linear regressions with accuracy as dependent variable.

	Model 1	Model 2
(Intercept)	2.375***	2.868***
	(0.136)	(0.354)
ACCHIGH	0.046	0.021
	(0.175)	(0.175)
ACCLOW	−0.209	−0.244
	(0.164)	(0.160)
TOURHIGH	−0.218	−0.274+
	(0.163)	(0.159)
TOURLOW	−0.236	−0.222
	(0.168)	(0.168)
Overestimation		0.220*
		(0.087)
Forecaster: Experimental economists		−0.907***
		(0.232)
Forecaster: Farm_Advisors_Croatia		−0.293
		(0.206)
Forecaster: Farm_Advisors_France		−0.713***
		(0.188)
Forecaster: Farm_Advisors_Italy		−0.066
		(0.206)
Forecaster: Experts_Spain		−0.426*
		(0.205)
Forecaster: Swedish_students		−0.592**
		(0.227)
Forecaster: Other		−0.600**
		(0.222)
Female		0.161
		(0.105)
Age		−0.004
		(0.005)
Background Agricultural Sciences/Farming		−0.008
		(0.145)
Background Other		−0.135
		(0.154)
Num. Obs.	4,488	4,408
R2	0.008	0.047
AIC	17,335.5	16,854.8
Prediction sample FE	Yes	Yes

Source: Own calculations. Notes: $+P < 0.1$, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. Clustered standard errors in parentheses.

Appendix 7. Additional linear regressions with accuracy as dependent variable.

	Model 1	Model 2	Model 3	Model 4
(Intercept)	2.4726***	2.4629***	2.4955***	2.4842***
	(0.1454)	(0.1462)	(0.1559)	(0.1569)
Dummy: Own country	0.1937*	0.2715+	0.2062**	0.2988+
	(0.0773)	(0.1638)	(0.0784)	(0.1666)
ACCHIGH	0.0076	0.0311	0.0076	0.0336
	(0.2019)	(0.2043)	(0.2021)	(0.2049)
ACCLOW	−0.1723	−0.1456	−0.1723	−0.1428
	(0.1894)	(0.1913)	(0.1896)	(0.1918)
TOURHIGH	−0.1884	−0.2082	−0.1884	−0.2059
	(0.1865)	(0.1902)	(0.1867)	(0.1908)
TOURLOW	−0.3942*	−0.3794+	−0.3942*	−0.3784+
	(0.1957)	(0.2001)	(0.1959)	(0.2006)
Dummy: Own country × ACCHIGH		−0.1880		−0.2081
		(0.2298)		(0.2315)
Dummy: Own country × ACCLOW		−0.2137		−0.2365
		(0.2430)		(0.2428)
Dummy: Own country × TOURHIGH		0.1581		0.1403
		(0.2510)		(0.2512)
Dummy: Own country × TOURLOW		−0.1188		−0.1266
		(0.2316)		(0.2331)
Controls				
Dummies indicating the predicted sample	No	No	Yes	Yes
Num. Obs.	3,328	3,328	3,328	3,328
R^2	0.008	0.009	0.011	0.011
R^2 Adj.	0.007	0.006	0.007	0.007
AIC	13,095.8	13,101.5	13,102.3	13,107.9
Std. errors	by: id	by: id	by: id	by: id

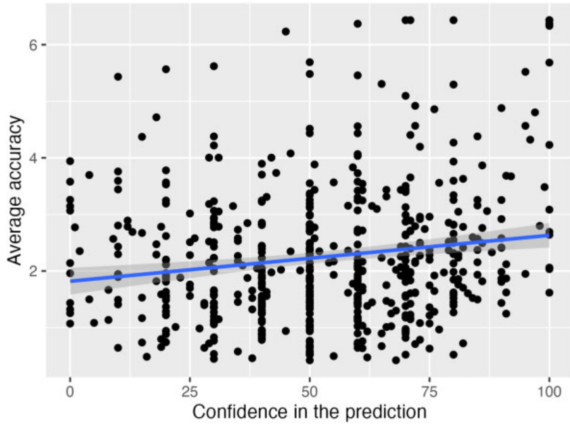
$+P < 0.1$, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

Appendix 8. Confidence in predictions by forecaster group.

Forecaster samples	Minimum	Median	Mean	SD	Maximum
Farm Advisors_Poland	0.00	60.00	60.16	21.70	100.00
Experimental economists	0.00	40.00	42.28	20.84	90.00
Farm Advisors_Croatia	12.00	70.00	64.93	17.36	99.00
Farm Advisors_France	1.00	40.00	41.04	20.52	81.00
Farm Advisors_Italy	0.00	66.00	60.98	27.06	100.00
Farm Advisors_Spain	0.00	52.00	50.88	22.02	91.00
Swedish students	0.00	50.00	50.81	25.23	100.00
Other	0.00	58.00	51.54	25.36	100.00

Note: Confidence on a scale from 0 (not confident at all) to 100 (very confident).

Appendix 9. Scatterplot of the participants confidence in their predictions and their average prediction accuracy ($n = 561$).



Notes: Own calculations; Blue line: linear fit.

References

Bellemare M. F., Lee Y. N. and Just D. R. (2020) 'Producer Attitudes Toward Output Price Risk: Experimental Evidence from the Lab and from the Field', *American Journal of Agricultural Economics*, 102: 806–25.

Binswanger H. P. (1980) 'Attitudes Toward Risk: Experimental Measurement in Rural India', *American Journal of Agricultural Economics*, 62: 395–407.

Bocquého G., Jacquet F. and Reynaud A. (2014) 'Expected Utility or Prospect Theory Maximisers? Assessing Farmers' risk Behaviour from Field-experiment Data', *European Review of Agricultural Economics*, 41: 135–72.

Bonjean I. (2023) 'Who Are the Loss-averse Farmers? Experimental Evidence from Structurally Estimated Risk Preferences', *European Review of Agricultural Economics*, 50: 421–56.

Britz W. (2022) 'Implementing an Approximation of Cumulative Prospect Theory into Mixed Linear Programming: an Application to Bio-Economic Modelling at Farm-Scale Considering Crop Insurance', *German Journal of Agricultural Economics*, 71: 184–203.

Camerer C. F. and Hogarth R. M. (1999) 'The Effects of Financial Incentives in Experiments: a Review and Capital-Labor-Production Framework', *Journal of Risk & Uncertainty*, 19: 7–42.

Cerroni S. (2020) 'Eliciting Farmers' subjective Probabilities, Risk, and Uncertainty Preferences Using Contextualized Field Experiments', *Agricultural Economics*, 51: 707–24.

Cerroni S., Bozzola M. and Rippo R. (2023) 'Eliciting Farmers' risk and Uncertainty Preferences Using Contextualized Field Experiments in Developing Countries: a Pilot study'. In: Mishra, A K, Kumbhakar, S, Lien, and G (eds.) *Managing Risk in Agriculture: a Development Perspective*, pp. 85–100. CABI: UK.

Cerroni S. et al. (2023) 'A Choice Matching Approach for Discrete Choice Analysis: an Experimental Investigation in the Laboratory', *Land Economics*, 99: 80–102.

Charness G., Gneezy U. and Imas A. (2013) 'Experimental Methods: Eliciting Risk Preferences', *Journal of Economic Behavior & Organization*, 87: 43–51.

Charness G., Gneezy U. and Rasocha V. (2021) 'Experimental Methods: Eliciting Beliefs', *Journal of Economic Behavior & Organization*, 189: 234–56.

Crosetto P. and Filippin A. (2016) 'A Theoretical and Experimental Appraisal of Four Risk Elicitation Methods', *Experimental Economics*, 19: 613–41.

Dave C. et al. (2010) 'Eliciting Risk Preferences: When Is Simple Better?' *Journal of Risk & Uncertainty*, 41: 219–43.

DellaVigna S., Otis N. and Vivalt E. (2020) 'Forecasting the Results of Experiments: Piloting an Elicitation Strategy', *AEA Papers and Proceedings*, 110: 75–9.

DellaVigna S. and Pope D. (2018a) 'Predicting Experimental Results: Who Knows What?' *Journal of Political Economy*, 126: 2410–56.

———. (2018b) 'What Motivates Effort? Evidence and Expert Forecasts', *The Review of Economic Studies*, 85: 1029–69.

DellaVigna S., Pope D. and Vivalt E. (2019) 'Predict Science to Improve Science', *Science*, 366: 428–9.

Dessart F. J., Barreiro-Hurlé J. and van Bavel R. (2019) 'Behavioural Factors Affecting the Adoption of Sustainable Farming Practices: a Policy-oriented Review', *European Review of Agricultural Economics*, 46: 417–71.

Drichoutis A. C. and Lusk J. L. (2016) 'What Can Multiple Price Lists Really Tell Us about Risk Preferences?' *Journal of Risk & Uncertainty*, 53: 89–106.

Eisele M., Troost C. and Berger T. (2021) 'How Bayesian Are Farmers When Making Climate Adaptation Decisions? A Computer Laboratory Experiment for Parameterising Models of Expectation Formation', *Journal of Agricultural Economics*, 72: 805–28.

Finger R. et al. (2024) 'A Note on European Farmers' preferences under Cumulative Prospect Theory', *Journal of Agricultural Economics*, 75: 465–72.

Finger R., Wüpper D. and McCallum C. (2023) 'The (in)Stability of Farmer Risk Preferences', *Journal of Agricultural Economics*, 74: 155–67.

Friedman D. et al. (2022) 'Varieties of Risk Preference Elicitation', *Games and Economic Behavior*, 133: 58–76.

Gigerenzer G. and Brighton H. (2009) 'Homo Heuristicus: Why Biased Minds Make Better Inferences', *Topics in Cognitive Science*, 1: 107–43.

Grossmann I. et al. and The Forecasting Collaborative. (2023) 'Insights into the Accuracy of Social Scientists' forecasts of Societal Change', *Nature Human Behaviour*, 7: 484–501.

Grüner S. et al. (2022) 'How (un)Informative Are Experiments with Students for Other Social Groups? A Study of Agricultural Students and Farmers', *Australian Journal of Agricultural and Resource Economics*, 66: 471–504.

Hardaker J. B. and Lien G. (2010) 'Probabilities for Decision Analysis in Agriculture and Rural Resource Economics: the Need for a Paradigm Change', *Agricultural Systems*, 103: 345–50.

Hellerstein D., Higgins N. and Horowitz J. (2013) 'The Predictive Power of Risk Preference Measures for Farming Decisions', *European Review of Agricultural Economics*, 40: 807–33.

Höhler J. et al. (2024) 'Perspectives on Stakeholder Participation in the Design of Economic Experiments for Agricultural Policymaking: Pros, Cons, and Twelve Recommendations for Researchers', *Applied Economic Perspectives and Policy*, 46: 338–59.

Huber R. et al. (2022) 'Bridging Behavioural Factors and Standard Bio-economic Modelling in an Agent-based Modelling Framework', *Journal of Agricultural Economics*, 73: 35–63.

Iyer P. et al. (2020) 'Measuring Farmer Risk Preferences in Europe: a Systematic Review', *Journal of Agricultural Economics*, 71: 3–26.

Maart-Noelck S. C. and Musshoff O. (2014) 'Measuring the Risk Attitude of Decision-makers: Are There Differences between Groups of Methods and Persons?' *Australian Journal of Agricultural and Resource Economics*, 58: 336–52.

Menapace L., Colson G. and Raffaelli R. (2016) 'A Comparison of Hypothetical Risk Attitude Elicitation Instruments for Explaining Farmer Crop Insurance Purchases', *European Review of Agricultural Economics*, 43: 113–35.

Milkman K. L. et al. (2022) 'A 680,000-person Megastudy of Nudges to Encourage Vaccination in Pharmacies', *Proceedings of the National Academy of Sciences*, 119: e2115126119.

Niederle M. and Vesterlund L. (2007) 'Do Women Shy Away from Competition? Do Men Compete Too Much?' *The Quarterly Journal of Economics*, 122: 1067–101.

Norris P. E. and Kramer R. A. (1990) 'The Elicitation of Subjective Probabilities with Applications in Agricultural Economics', *Review of Marketing and Agricultural Economics*, 58: 127–47.

Palm-Forster L. H., Suter J. F. and Messer K. D. (2019) 'Experimental Evidence on Policy Approaches That Link Agricultural Subsidies to Water Quality Outcomes', *American Journal of Agricultural Economics*, 101: 109–33.

Rakotonarivo O. S., Schaafsma M. and Hockley N. (2016) 'A Systematic Review of the Reliability and Validity of Discrete Choice Experiments in Valuing Non-market Environmental Goods', *Journal of Environmental Management*, 183: 98–109.

Rigby R. A. and Stasinopoulos D. M. (2005) 'Generalized Additive Models for Location, Scale and Shape (with discussion) ', *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54: 507–54.

Rommel J. et al. (2023a) 'Farmers Risk Preferences in 11 European Farming Systems: a Multi-country Replication of Bocquého et al. (2014)', *Applied Economic Perspectives and Policy*, 45: 1374–99.
———. (2023b) 'Learning about German Farmers' willingness to Cooperate from Public Goods Games and Expert Predictions', *Q Open*, 3: qoac023.
Ruhinduka R. D. et al. (2020) 'Smallholder Rice Farmers' post-harvest Decisions: Preferences and Structural Factors', *European Review of Agricultural Economics*, 47: 1587–620.
Rust N. A. et al. (2022) 'Have Farmers Had Enough of Experts?' *Environmental Management*, 69: 31–44.
Schaak H. et al. (2017) 'The Predictive Power of Experimental Risk Attitude Measures for Farm Diversification', *Schriften der Gesellschaft für Wirtschafts- und Sozialwissenschaften des Landbaues*, 52: 87–97.
Stasinopoulos D. M. and Rigby R. A. (2007) 'Generalized Additive Models for Location Scale and Shape (GAMLSS) in R', *Journal of Statistical Software*, 23: 1–46.
Tanaka T., Camerer C. F. and Nguyen Q. (2010) 'Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam', *American Economic Review*, 100: 557–71.
Trautmann S. T. and van de Kuilen G. (2015). 'Belief Elicitation: a Horse Race among Truth Serums', *The Economic Journal*, 125: 2116–35.
Vivalt E. and Coville A. (2023) 'How Do Policymakers Update Their Beliefs?' *Journal of Development Economics*, 165: 103121.
Vollmer E., Hermann D. and Mußhoff O. (2017) 'Is the Risk Attitude Measured with the Holt and Laury Task Reflected in Farmers' production Risk?' *European Review of Agricultural Economics*, 44: 399–424.
Voslinsky A. and Azar O. H. (2021) 'Incentives in Experimental Economics', *Journal of Behavioral and Experimental Economics*, 93: 101706.
Wuepper D., Roleff N. and Finger R. (2021) 'Does It Matter Who Advises Farmers? Pest Management Choices with Public and Private Extension', *Food Policy*, 99: 101995.
Yu C. W., Zhang Y. J. and Zuo S. X. (2021) 'Multiple Switching and Data Quality in the Multiple Price List', *The Review of Economics and Statistics*, 103: 136–50.