The Psychometric Properties of Probability and Quantile Forecasts

Sophie Ma Zhu[1,2] Nikolay Petrov[3,2] David Budescu[4,2] Ezra Karger[2,5] Mark Himmelstein[6,2]

[1]University of British Columbia [2]Forecasting Research Institute [3]University of Cambridge
[4]Fordham University [5]Federal Reserve Bank of Chicago [6]Georgia Institute of Technology

INTRODUCTION

In forecasting tournaments, it has been common to estimate the probabilities of discrete events. For predictions of continuous quantities, the possible range of outcome values is divided into exclusive bins so that probabilities for each bin can be elicited (i.e., ***probability format***).

An alternative approach is to directly elicit forecasts about quantiles of the continuous quantity (i.e., ***quantile format***).

A significant gap still exists in understanding the quantile format or comparisons between the two approaches.

METHODS

Participants: 1,147 respondents recruited online (64.5% female), aged 17-84 years (M=43.4, SD=13.2).

Forecasting tasks: Five unique forms (A-E), each with six forecasting questions on continuous quantities across various topics, selected from an existing item pool based on previous tournaments.

Design: 7-survey longitudinal study, with five surveys focused on forecasting tasks. Participants were randomly assigned to form sequences using a Latin square design. Each participant completed the same two forms in Surveys 1 & 5, in reversed formats, and the remaining three forms in Surveys 2-4.

We used simulated data to better understand how scores behave in both formats. (Figure 3)
*- When items are resolved **in an expected way** (close to 0), the range of probability forecast scores is truncated.*
*- When items are resolved **in an unexpected way** (far away from 0), only poor forecasters who assigned extreme probabilities could get lucky and receive a perfect score.*

RESULTS

Table 1. Internal consistency (Cronbach's α) for five forms of forecasting questions across both formats

Form	Survey 1		Survey 5	
	Probability format	Quantile format	Probability format	Quantile format
A	.06	.66	.11	.73
B	.45	.54	.55	.68
C	.50	.60	.53	.66
D	.25	.71	.20	.66
E	.42	.66	.56	.76

Cronbach's α values were higher and more consistent across different forms in the ***quantile format***. Acceptable reliability is achievable with as few as six items.

The correlation between person-level accuracy scores in Surveys 1 and 5 was .26 for ***probability forecasts*** and .58 for ***quantile forecasts.***

When using the accuracy scores from the three interim surveys (in-sample) to predict accuracy in Surveys 1 & 5 (out-of-sample), variability in the accuracy of ***quantile forecasts*** was more statistically explainable (R^2 = .66) compared to ***probability forecasts*** (R^2 = .17).

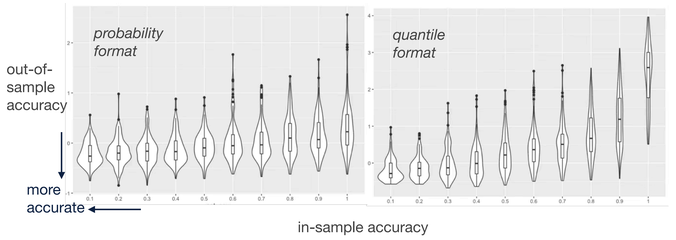

Figure 1. Distribution of out-of-sample accuracy scores grouped by deciles of in-sample accuracy

Figure 1 illustrates that participants who were more accurate in the interim surveys clustered at the lower end of out-of-sample accuracy scores, while less accurate participants showed wider variation.

Overall, median out-of-sample accuracy for ***quantile forecasts*** decreased steadily across deciles of in-sample accuracy, indicating a clear relationship. In contrast, in-sample accuracy was less predictive of out-of-sample performance for ***probability forecasts***, which showed a more uniform distribution.

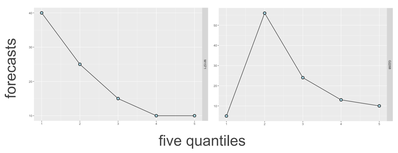

Figure 2. Examples of incoherent forecasts in responses to the quantile questions

For some quantile forecasts, the five numeric inputs were not monotonically increasing (see examples in Figure 2). Reversed order and bell-curve shape suggested confusion and comprehension challenges.

What will be the closing value for the U.S. Dollar against the Russian Ruble on May 31, 2024?

Probability Format - with values fixed, accuracy measured by the ordinal Brier score

Remember, your probabilities must total up to 100.

	0 10 20 30 40 50 60 70 80 90 100	
Less than 80		0
Greater than or equal to 80 and less than 85		0
Greater than or equal to 85 and less than 90		0
Greater than or equal to 90 and less than 95		0
Greater than or equal to 95		0
Total:		0

Quantile Format - with probabilities fixed, accuracy measured by the S-score

For each percentage listed below, your answer should indicate you believe there is that percent chance the outcome will be ____ or less.

5%	
25%	
50%	
75%	
95%	

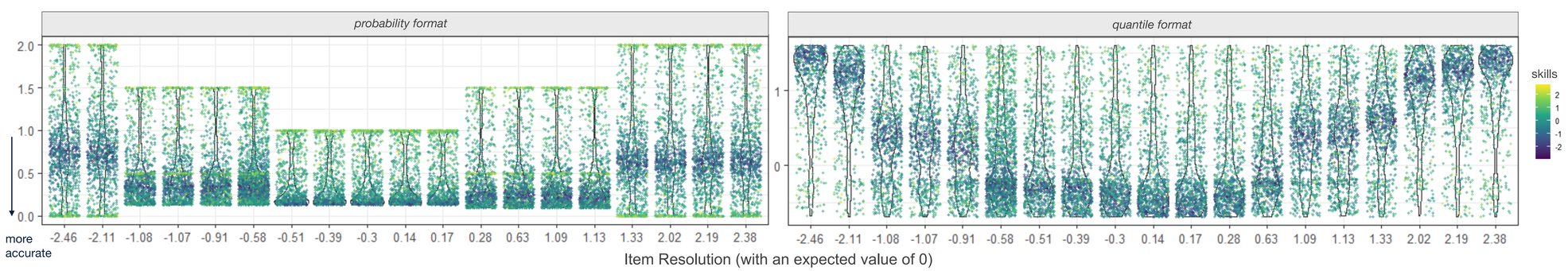

Figure 3. Simulated data: Accuracy score distribution for questions with different difficulty levels, colored by individual forecasting skills (i.e., average amount of errors)

CONCLUSION

Our results provide strong evidence that eliciting uncertainty in forecasting via quantile (fixed probability) format has a reliability advantage over the more common probability (fixed variable) format. Practitioners and researchers are advised to utilize this format especially when their goal is to spot high-performing forecasting talent.

Improvement is needed in the precise elicitation mechanisms and user interface to help facilitate comprehension and prevent errors.

Questions/comments: *zhuma@student.ubc.ca*

Society for Judgment and Decision Making Annual Meeting 2024