Wissenschaftszentrum Berlin
für Sozialforschung

Sabine Kröger
Thibaud Pierrot

What point of a distribution summarises point predictions?

Discussion Paper

SP II 2019–212

October 2019

WZB Berlin Social Science Center

Research Area
Markets and Choice

Research Unit
Market Behavior

Wissenschaftszentrum Berlin für Sozialforschung gGmbH
Reichpietschufer 50
10785 Berlin
Germany
www.wzb.eu

Sabine Kröger, Thibaud Pierrot
What point of a distribution summarises point predictions?

Affiliation of the authors:

Sabine Kröger
Laval University

Thibaud Pierrot
WZB Berlin Social Science Center and Technische Universität Berlin

Wissenschaftszentrum Berlin für Sozialforschung gGmbH
Reichpietschufer 50
10785 Berlin
Germany
www.wzb.eu

Abstract

What point of a distribution summarises point predictions?

by Sabine Kröger and Thibaud Pierrot*

In this article, we study the point predictions that forecasters report when they are asked to predict the realisation of an iid random variable. We set up a laboratory experiment where the participants act as forecasters predicting the next realisation of random draws coming from different "objectively known" distributions which vary in the location of their central tendencies. As is standard in survey measures, the subjects in our experiment must report their best guess of the next draw as a forecast. We find that most of the forecasters report point predictions that are close to one of the three main central tendencies (mean, median or mode) of the distributions provided, with a majority corresponding to the mode. Our analysis also shows that when selecting a point prediction, people have in mind a numerical value (e.g. the mean or the mode) rather than a specific percentile of the underlying distribution. Only 5% of the forecasts reported during the experiment are based on a percentile while almost 60% are based on a numerical value.

Keywords: subjective expectations, forecasting, eliciting point predictions, experiment

JEL classification C91; C72; D84

* E-mail: sabine.kroger@ecn.ulaval.ca, thib.pierrot@gmail.com.

1 Introduction

Subjective expectations data provide important information regarding individual and expert opinions as well as their uncertainty about the future. Such data have become a cornerstone in economic analysis, feeding into both macro-econometric forecasting models and micro-econometric models of individual decision-making. They allow us to separate preferences from expectations without relying on the rational expectations hypothesis (Manski, 2002, 2004).

In this paper, we study how the forecasts produced by individuals reflect their subjective expectations. We explore the relationship between these forecasts and the probability distributions of the random variables of interest. Our goal is to pin down how people summarise an entire distribution of probabilities into a single point when they are asked for their best guess regarding the future realisation of a random variable.

Economists consider uncertainty in terms of (non-degenerate) outcome distributions of an event and those who use expectations data presume that respondents have a unique subjective distribution in mind.[1] However, with some exceptions (Delavande 2014), most surveys (regarding professionals or households) and experimental studies do not elicit subjective distributions but ask for forecasts. Thus, the respondents are supposed to report their beliefs in the form of a single point. Without any knowledge regarding the individuals' forecasting rules, it is impossible to compare and use the collected points. Therefore, the researchers presume that the forecasts are selected using a homogeneous rule, i.e. that each individual selects a point that matches the same characteristic of the distribution from which it was chosen.

In the literature, the standard assumption is to consider that the forecast corresponds to the mean of the unique distribution as a result of applying a quadratic, and thus symmetric and unbiased, loss function to the distance between this forecast and the possible realisations of the random variable. Many interpretations of empirical results rely on this assumption. For example, the dispersion

[1] Even multiple prior models, where the decision-maker faces (Knightian) uncertainty, reduce uncertainty to one subjective distribution before the decision-maker chooses between alternatives.

of individual forecasts is often perceived as a measure for the severity of uncertainty and of differences in available information between individuals (Pesaran and Weale, 2006). However, several researchers have pointed out that part of the dispersion empirically measured can be accounted for by relaxing the assumption of forecasters using the same forecasting rule (Manski, 1999; Das, Dominitz & van Soest, 1999).

The forecasting rules have been studied empirically using the Survey of Professional Forecasters (hereafter SPF). This dataset gathers the predictions of the US GDP growth and inflation made by professional forecasters.[2] Elliott, Komunjer & Timmermann (2008) suggest that an asymmetric forecasting loss function, varying across forecasters, would rationalise the predictions reported in the SPF. Engelberg, Manski & Williams (2009)(EMW hereafter) compare the elicited point predictions with their interval-elicited subjective probability distributions. They find that the forecasting rules vary among the experts. They specifically stress the joint effect of belief dispersion and forecasting rule heterogeneity.[3] Their findings highlight the difficulty in interpreting and using the point predictions, e.g. in models of choice under uncertainty or when interpreting the dispersion of forecasters' individual point predictions.

In this work, we extend the analysis of EMW by conducting a forecast experiment in the laboratory using objectively known distributions instead of elicited subjective distributions. This design allows us to specifically create distributions with dispersed central tendencies. Contrarily to most of the probability distributions elicited in the SPF, ours are not Gaussian shaped and therefore do not concentrate their mean, median and mode within the same interval. Hence, we can investigate the influence of each of these central tendencies separately on the forecast selected to summarise the probability distribution.

Moreover, our set of distributions contains pairs of mirror images. Comparing the points reported for these pairs enables us to test whether our subjects' forecasting rules are based on a numerical value or a percentile of the underlying distribution. Indeed, some models of decision-making under risk, such as the quantile utility

[2]Federal Reserve Bank of Philadelphia at http://www.phil.org/econ/spf/spfbib.html

[3]Reporting rules are heterogeneous but mostly consistent with one of the three central tendencies of the distribution, the mode, the median and the mean.

model of Manski (1988), postulate that when choosing among multiple risky bets, a decision-maker may maximise a particular percentile of the probability distribution rather than its mean as in the expected utility models. Our design allows us to look at the empirical relevance of this class of models by investigating whether the predictions made by individuals are related to a specific percentile of the distributions displayed or to a numerically-based characteristic such as the means of these distributions.

Finally, our design insures control over the information possessed by the forecasters and over the way they produce their prediction, i.e. alone, in groups, with friends, family, colleagues or with the support of prediction software. In our experiment, all the participants had the same information on the data generating process in the form of histograms of the objective distribution. They processed this information alone and made their forecasts by themselves.

In line with the empirical findings of EMW (2009), we find that most of the forecasts reported are close to a central tendency of the distribution: either its mean, median or mode. Among those forecasts the majority correspond to the mode of the distribution. This result indicates that, without further instructions, when they choose a forecast most individuals maximise the probability of being exactly right.

The subjects' forecasting rules are mostly based on the selection of a numerical value of the probability distribution (e.g. the mean or the mode) rather than of a percentile (e.g. the 60th or 40th percentile). Nevertheless, we observe that the characteristics of both the forecasters and the probability distributions have a significant impact on the rule applied to selecting a point prediction. This finding highlights the potential issues regarding the interpretation of forecasts reported in surveys as the subjective expectations of the respondents.

The paper is organised as follows: Our experiment is introduced in section 2, the results are presented in section 3. We close with a discussion in section 4.

2 Experiment

We elicited point predictions from objectively known distributions within a computerised experiment (using software z-Tree, Fischbacher, 2007) that was conducted at the LEEL (Laboratory of Economic Experiments at Laval University, Quebec, Canada). In the experiment, the participants were given the role of professional forecasters whose task it was to predict the future profits of 10 different companies. They were provided with the profit distribution of each company in the form of a histogram with eight possible profit realisations: $\{-35, -25, -15, -5, 5, 15, 25, 35\}$.

While entering their point predictions, the subjects knew that the computer would draw the profit realisation that they had to forecast from the distribution that was displayed on their screen. As is standard in survey practice, we did not give the participants further instructions regarding which point of the distribution to report. We asked them to provide "their best guess" regarding the next realisation of the profit for the company and specified that their forecast must be "as close as possible" to this profit. We also deliberately decided against using proper scoring rules to incentivise the forecasts' accuracy as monetary incentives are hardly ever used in surveys.[4]

The predictions reported could be any integer between the lower and upper bound of the profit support. It was not mandatory to choose a forecast corresponding to one of the eight possible realisations. The subjects made 10 consecutive predictions for each of the ten companies. After each forecast, the computer drew the profit realisation and displayed it with all the past predictions and realisations in a history table. Once 10 forecasts were made and 10 realisations were drawn for a company, each participant started over again with a new company that had a different profit distribution and an empty history table. Here, we provide an example of a participant's screen in Figure 1.

[4]The proper scoring rules have the property to maximise participants' expected payoffs when they report particular points of the distributions - e.g. a quadratic scoring rule maximises the expected gain when the mean is reported as a forecast. To properly answer our research question, we have to avoid such incentives. Indeed, we are documenting here how individuals make forecasts in real life or report predictions during surveys - which for practical reasons do not use scoring rules.

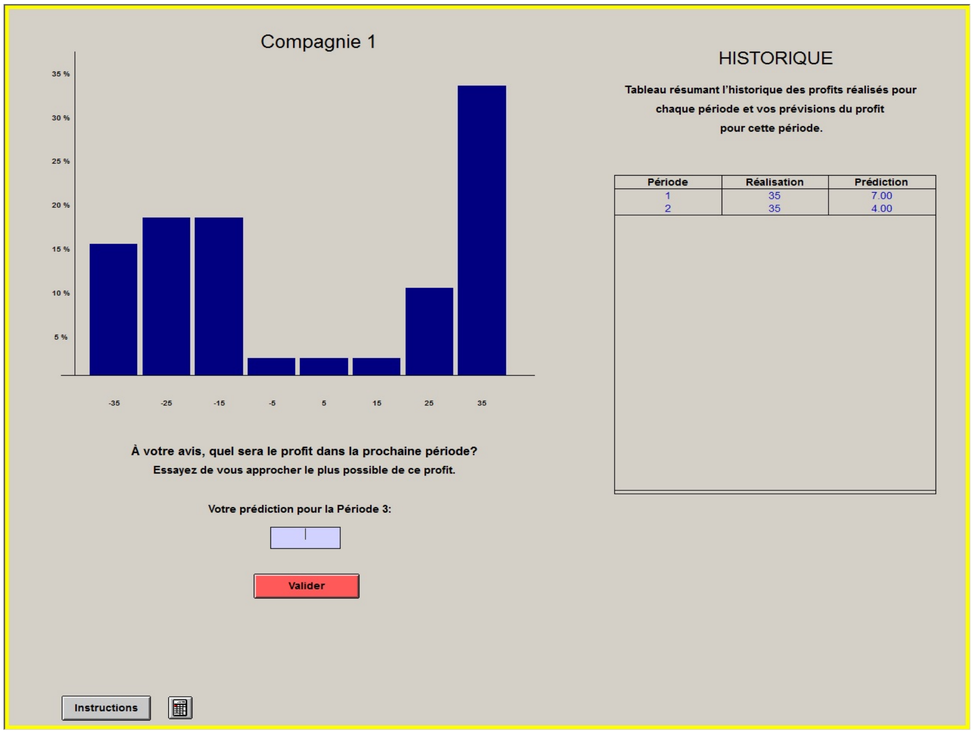

Figure 1: Screenshot of prediction task.

Every participant was asked to make forecasts for 10 different companies, for a total of one hundred predictions. The companies appeared in a random order during the experiment and were represented by a unique profit distribution. These distributions are presented in Figure 2. They were constructed in pairs containing each a *standard* distribution and its *image* counterpart. In each pair the *image* is the mirror image of the *standard*. It was obtained by multiplying the profits of the *standard* distribution by '-1'. This feature of our design allows us to investigate whether our subjects' forecasting rules are based on a percentile or a numerical value of the distributions. Appendix B explains in greater detail how we constructed the distributions.

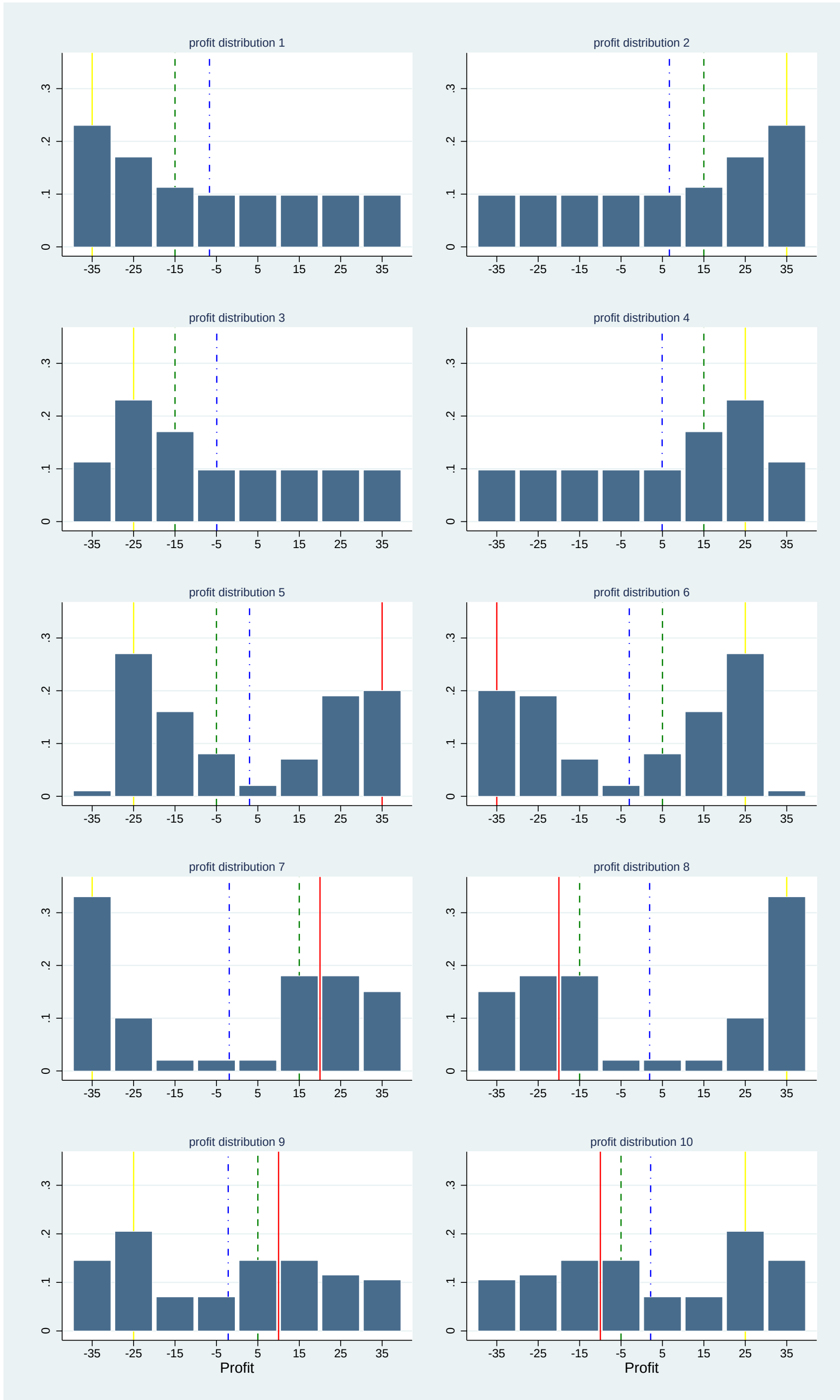

Figure 2: Ten profit distributions that participants saw in random order and for which they predicted the next or the next six draws.
Distributions 1, 3, 5, 7 and 9 are the *standard* distributions while distributions 2, 4, 6, 8 and 10 are their *image* counterpart, i.e. all profits are multiplied by -1, e.g. distribution 1 has the mode at -35, distribution 2 has the mode at 35.
—— —— Full lines indicate the mode (yellow) or the second highest mode (red).
- - - - Dashed lines (green) indicate the median.
-.-.-.- Dashed-dotted lines (blue) indicate the mean.

The five experimental sessions were conducted in September 2014 and 2016.[5] In total 41 participants (29 in 2014 and 12 in 2016) took part in the experiment. A session lasted between 1h40 and 2h15 in total. At the beginning the instructions were displayed on the screen in a short video. They were also accessible in a written version during the whole experiment. Then the subjects performed the forecasting task and, at the end of the experiment, they answered a post-experimental questionnaire in which we collected some background information on age, gender and level of education.

This questionnaire also contained the Berlin Numeracy Test (BNT) that measures the ability to represent, store and accurately process mathematical operations.[6] The BNT is a psychometric instrument that assesses and represents the statistical and risk literacy of a person on a scale of 1 to 4. A value of 1 in the BNT presents the lowest level of numeracy and 4 the highest. The BNT is particularly powerful when measuring the cognitive ability of individuals to understand and manipulate ratio concepts, proportions, probabilities and percentages. It was designed in order to achieve a high level of psychometric discriminability among highly-educated individuals (e.g. college students and graduates, medical professionals).

We recruited the participants among the students and personnel of Laval University using an online system. They received a 5 CAD show-up fee upon arrival at the laboratory, a fixed amount of 20 CAD (30 CAD in 2016) for the completion of the prediction task and another 5 CAD for answering the post-experimental questionnaire.

[5]In the 2016 sessions, multiple treatments of the experiment were conducted. In one of them, the first half of the experiment was a replication of the design conducted in 2014 and described here. The second was a treatment varying the elicitation questions that were asked. In this paper, we pool the data collected in 2014 with its 2016 replication to increase our sample size. No significant differences were found between the two samples, both in terms of the individual characteristics of the participants (age, gender, numeracy) and in terms of their behaviour during the experiment (forecasts reported, time).

We restricted the number of companies in the 2016 sessions to 10 for the experiment to have a reasonable length, whereas participants in 2014 were asked to make predictions for a total of 12 companies, i.e. one additional *object* and its *image* distribution. We exclude the analysis of those two distributions from the 2014 data to make both data sets comparable. The other treatments conducted in 2016 are analysed in a companion paper (Kröger & Pierrot, 2019).

[6]We employed the computerised version of the BNT which is adaptive, i.e. follow-up questions depend on previous questions, and asks 2 to 3 out of 4 questions.

3 Results

We aim to characterise the relationship between the probability distribution of a random variable and the forecast selected by an individual to predict the next realization of this variable. To investigate this relationship it is absolutely crucial that the predictions used for our analysis are not affected by previous realisations of profits. Hence, we exclusively use the first prediction reported for each profit distribution. This restriction insures the independence of the forecasts analysed here. It leaves us with 410 predictions made by 41 subjects

3.1 Descriptive Statistics

In Table 1 we report the background characteristics of the participants. A session lasted an average of one hour and 13 minutes. The subjects were 30 years old, on average, about 30% of them were women and 66% held an academic degree. Their score in the numeracy test was 2.4, on average (on a 1 to 4 scale), and they took an average of 30 seconds to produce their first forecast.

Gender	0.32		=1 if female
Age	30.4	(9.0)	Age of the subjects
Education	0.66		=1 when holding an academic degree
Numeracy	2.4	(1.2)	Numeracy measured by the BNT on a scale for 1 (Worst) to 4 (Best)
Time per Prediction	$0.5'$	$(1.0')$	Time taken by the subjects to produce one forecast (in minutes)
Session Duration	$73.6'$	$(22.0')$	Time taken to complete the session (in minutes)
# of subjects	41		

Table 1: Measures of Background Characteristics and Time Taken for the Experiment (standard deviations in parentheses)

3.2 Forecasts Reported

We start our investigation by breaking down each forecast $y_{i,d}$ produced by an individual i for a distribution d into two core elements: (1) its numerical value ν_y and (2) the percentile of the underlying distribution to which it corresponds $\tau_{y,d}$. For example, if the prediction of subject 1 for the distribution 1 is 15 (i.e. $y_{1,1} = 15$), then its numerical value, ν_y, is also 15 and its associated percentile, $\tau_{y,d}$, is 50. Indeed, for distribution 1 a profit of 15 corresponds to the 50th percentile of the distribution. For another distribution it would be associated with a different percentile.

Separating a point prediction into its numerical value and its corresponding percentile allows us to consider value-based and percentile-based forecasting rules in our investigation. We display in the first section of Table 2 the average forecast $\overline{y}$ reported by our subjects for each distribution d as well as its corresponding percentile $\overline{\tau_{y,d}}$.

	Distribution (d)										
	1	2	3	4	5	6	7	8	9	10	Average
$\overline{y}$ $(= \overline{\nu_y})$	-16.8	22.9	-12.1	15.2	-11.6	9.9	-16.1	20	-8.3	9.9	1.3
σ_y	18.6	16.0	16.6	17.3	17.2	22.0	24.9	23.6	16.1	16.1	24.1
$\overline{\tau_{y,d}}$	34	75	42	63	35	66	26	78	43	59	52
σ_τ	29	27	22	23	23	27	33	32	18	18	31
$\overline{\tau_{y,d}} = 50$ (p-value)	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.18
	proportion of y close to a central tendency (ct)										
$\overline{y} = ct$	0.61	0.76	0.78	0.78	0.87	0.68	0.73	0.73	0.82	0.82	0.76
$\overline{y} = [mean \vert \overline{y} = ct]$	0.20	0.13	0.9	0.6	0.19	0.7	0.16	0.7	0.9	0.21	0.13
$\overline{y} = [median \vert \overline{y} = ct]$	0.20	0.26	0.25	0.31	0.17	0.7	0.6	0.7	0.38	0.23	0.20
$\overline{y} = [mode \vert \overline{y} = ct]$	0.60	0.61	0.66	0.63	0.64	0.86	0.77	0.87	0.53	0.56	0.67
	distance as a proportion of the support										
$\vert \overline{y} - mean \vert$	0.27	0.29	0.23	0.26	0.28	0.34	0.38	0.40	0.22	0.22	0.29
$\vert \overline{y} - median \vert$	0.23	0.21	0.18	0.18	0.22	0.29	0.50	0.55	0.23	0.26	0.29
$\vert \overline{y} - mode \vert$	0.26	0.17	0.18	0.15	0.19	0.22	0.27	0.21	0.24	0.22	0.21

Table 2: Average predictions with their corresponding percentiles and measures of proximity with the central tendencies of the underlying distributions.

3.2.1 Percentiles selected

The mean percentile reported for all the distributions combined is 52.1 which is not significantly different from the median according to a two-sided t-test including fixed effects for the individuals ($p = 0.18$). Nevertheless, we observe substantial variations across the different distributions indicating that most of the subjects do not consistently report a particular percentile of a distribution as their point forecast. For example, on average, the forecasters report the 34th percentile for the first distribution ($\overline{\tau_{y,1}} = 34$) and the 75th for the second distribution ($\overline{\tau_{y,2}} = 75$).

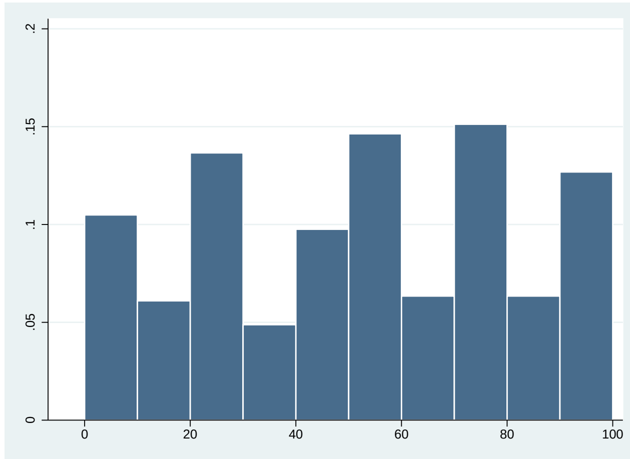

Figure 3: Relative frequencies of percentiles corresponding to reported forecasts

Moreover, the standard deviations associated with the percentiles reported are rather large, both for each distribution (between $\sigma_\tau = 17.5$ and $\sigma_\tau = 32.8$) and for the entire set of distributions ($\sigma_\tau = 30.8$) which confirms that the forecasts reported are unlikely to be based on a specific percentile. To illustrate this finding, we show the frequencies of each percentile reported during the experiment in Figure 2. The forecasters report many different percentiles, none of which seem to be more popular than the others.

Yet we observe an interesting pattern related to the symmetric properties of our distributions. The results presented in Table 2 show that, on average, the forecasts selected by our subjects correspond to a percentile that is below the median for the distributions with a mode below zero and above the median otherwise. This observation indicates that the mode is likely to be a key factor driving the choice of point predictions.

3.2.2 Relationship between forecasts and central tendencies

We now consider the correspondence between the predictions selected by our subjects to forecast the next realisation of profit and the three main central tendencies of the probability distributions, i.e. the mean, the median and the mode. This analysis follows EMW (2009). We replicate their findings and add the complementary analyses that our experimental setting permits.

In the second section of Table 2 we display the proportion of forecasts that are close to at least one of the three tendencies, as well as the proportions of these forecasts that are close to each tendency taken separately. We classify a prediction as close to the central tendency of a distribution if this prediction and this tendency fall into the same bin, where a bin refers to the predefined bound of the displayed histogram ($[-35; -20]$, $[-20; -10]$, $[-10; -0]$..., $[20; 35]$).

In all the distributions, we find that an average of 76% of the predictions correspond to one of the three main central tendencies (mean, median or mode). Thirteen percent of them correspond to the mean, 20% to the median and 67% to the mode. The modal response is therefore the most prominent one among our subjects. Although we witness variations across the different distributions, this observation holds for each individual distribution.

As a complementary measure, the third section of Table 2 contains the average absolute distance between each of the three tendencies and the point predictions. These distances confirm that the forecasts are close to the mode of the underlying distribution, but for some distributions the average distance to the median is very similar or even outperforms the distance to the mode.

To sum up, our results are in line with the findings of EMW (2009). They reveal that most of the subjects report point predictions that correspond to the mode of the probability distribution. Seventy-six percent of the point forecasts are close to one tendency, a slightly lower proportion than the 83% observed in the Survey of Professional Forecasters. This difference may have multiple explanations. First, the distributions in our experiment were objective and chosen by the experimenter while, in the SPF, the distributions were declared by the forecasters themselves and therefore subject to measurement errors. Second, in the dataset of EMW the three tendencies, carefully separated in our distributions, consistently lay in the same interval.

3.3 Do subjects report a numerical value or a percentile of the distribution?

Economic theory acknowledges that a forecaster asked to summarise a probability distribution by a single point may report either a particular characteristic of the distribution, i.e. its mean or mode, or a specific percentile of this distribution. For example, in the context of decision-making, Expected Utility Theory (Von Neumann and Morgenstern,1944 ; Savage, 1971) presumes that people use the mean of their utility distributions as point of reference, whereas in ordinal utility models (Manski, 1988) decision-makers use percentiles of the distributions to inform their decision.

Even though we cannot provide a definitive test here of the respective validity of these theories, our experimental design allows us to investigate the consistency of our subjects' forecasting rule with regard to either a percentile or a numerical characteristic of a given distribution. This analysis can be performed by exploiting the fact that the subjects participating in our experiment were confronted with five pairs of two identical distributions that are the mirror image of each other. By comparing the two forecasts reported for each pair, we can pin down whether a forecaster is more likely to base his prediction on a numerical value or a percentile of the displayed distribution as his forecast.

For example, let us consider a person who summarises the first distribution by reporting -6.7 as his forecast. This forecast is the mean of the first distribution and it is its 54^{th} percentile as well. To better understand what made the subject select -6.7 to predict the next realisation of the random variable, we look at the forecast that he chooses for the second distribution, which is the mirror image of the first. If the subject's forecast for this distribution is $+6.7$, which corresponds to the 46^{th} percentile, we consider this subject to be reporting a point based on a numerical characteristic of the distribution. On the other hand, if he reports 15, which correspond to the 54^{th} percentile, the subject is classified as reporting forecasts based on a percentile of the underlying distribution.

Comparing the reported summary statistics and corresponding percentiles of both distributions to one another allows us to better understand whether a person is basing his forecast on a numerical value and/or a percentile of the underlying distribution and how much the forecasts in a pair deviate from both.

3.3.1 Deviations from the points and percentiles reported in a pair of distributions : the α and β measures

We introduce two variables measuring the bias of a distribution summary that a person reports compared to a point of the distribution or a percentile. More precisely, we define the parameters

$$\alpha = \frac{1}{2}\left(1 + \frac{y_{-d}}{y_d}\right) \tag{1}$$

as the ratio of the point predictions for the reference distribution y_d and the one from its mirror image y_{-d} and equivalently

$$\beta = 1 - \frac{\tau_{d|y_d}}{\tau_{-d|y_{-d}}} \tag{2}$$

as the ratio of the percentile corresponding to the prediction for the reference distribution d and the percentile corresponding to the prediction of its mirror image $-d$. α measures the bias in terms of numerical value (i.e. of quantity or outcome) and β in terms of percentiles. For computing the two bias parameters, we exploit two convenient facts. First, the distributions to summarise have a symmetric support, and second, we have some degree of freedom to choose which of the distributions in a pair is the reference (*object*) distribution.

We arrange the distributions in such a way that we always divide by the larger of the two numbers. Hence, for every pair of distributions, α and β are in the interval $[0, 1]$. An α of zero implies that the two forecasts reported are the exact opposite numerical value of each other, i.e. $y_d = -1 \cdot y_{-d}$. This occurs, for example, when a subject consistently reports the mode or the mean of both distributions in a pair. On the contrary, $\alpha = 1$ when the same point is reported for both distributions, i.e. $y_d = y_{-d}$. Note that when a subject consistently reports the median, both biases are equal to zero, $\alpha = 0$ and $\beta = 0$. However, the more the reported percentile is located in one of the extremes, the larger the bias for α. The same applies vice versa. For a person basing their prediction on a numerical value, the further it is from the median, the larger the bias for β.

In our example above, a person who reports the mean value for both distributions of a pair would have a value of $\alpha = (1 + (-6.7/6.7))/2 = 0$ and $\beta = 1 - 46/54 = 0.25$. A person who predicts the same percentile as for the first distribution, would have a deviation of $\beta = 1 - 0.54/0.54 = 0$ for the percentile measure and a deviation of $\alpha = (1 + (-6.7/15))/2 = 0.28$ for the point. In this example, the reference distribution to compute α was distribution 2 and the one for β was distribution 1.

3.3.2 Descriptive statistics on α and β measures

Figure 4 shows the distribution of α and β obtained from the experimental forecasts in two histograms. For almost 60% of the forecast's pairs, α is close to zero, implying that $y_d = -1 \cdot y_{-d}$ for most of the predictions reported. The average α is 0.30 and its median is 0.14. Comparatively, we observe much higher values for β which has an average value of 0.55 and a median of 0.56. Overall, less than 10% of the β computed are close to zero.

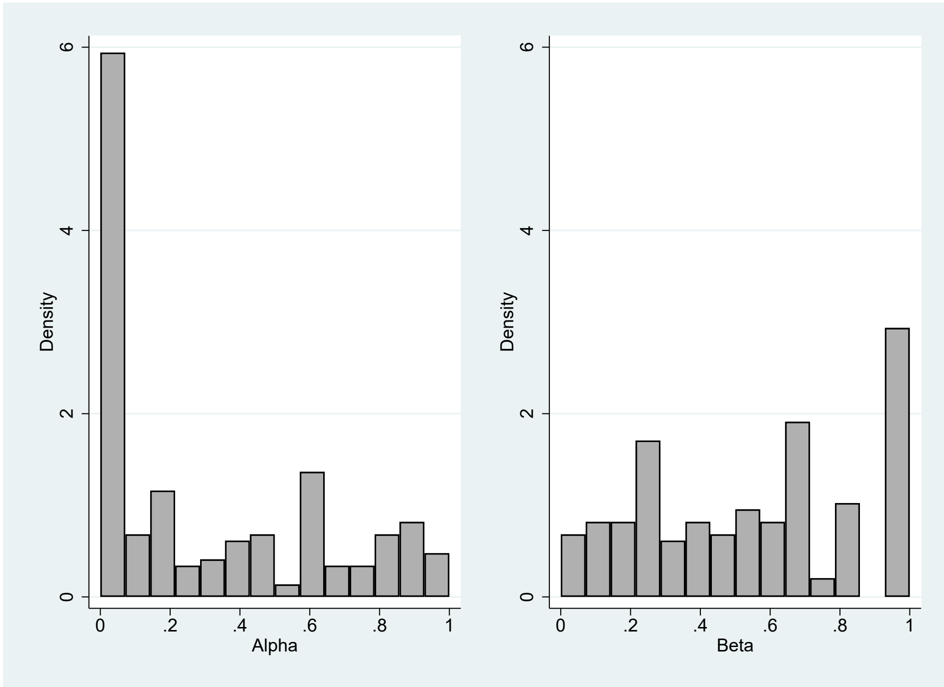

Figure 4: Distribution of α and β estimates.

These results indicate that almost 60% of the predictions reported are consistent with a rule based on forecasting a numerical value of the distribution. In contrast, very few subjects seem to consistently select a particular percentile of the distribution as their forecast. Among the 60% of forecasts with an $\alpha <= 0.1$, we find than 72% are close to the mode, 8% are close to the median and only 5% are close to the mean. These findings are in line with our previous results and further emphasize the preeminence of the modal report. They also demonstrate the heterogeneity of forecasting rules between forecasters in our sample. Indeed, even if forecasting the mode is a popular rule, many forecasts are neither close nor consistent with this rule.

4 Summary and Discussion

In this work we present an experimental study on what points people report when they are asked to predict the realisation of an *iid* random variable with a known distribution. In line with the empirical findings of EMW (2009), we find that most of our subjects report a central tendency of the distribution as their forecast. Of the forecasts selected, 76% correspond to either the mean, the median or the mode of the underlying distribution.

Our experimental design also allows us to derive further insights on the relationship between point predictions and probability distributions. We can discriminate among the three main tendencies and show that the most reported one is the mode. Of the forecasts that correspond to one of the three tendencies, 67% are close to the mode of the distribution. This result indicates that, without further instructions, most individuals tend to maximise the probability of being exactly right when they choose a forecast.

Moreover, thanks to the control that we exert over the distributions presented to the participants, we can investigate how the forecasts are selected by the subjects. Our results indicate that the subjects' forecasting rules are mostly based on a numerical value of the probability distribution rather than on a percentile. Only 5% of the forecasts chosen are consistent with a forecasting rule based on a percentile compared to almost 60% for a rule based on a numerical value. This finding is in line with the fact that a majority of the subjects choose the mode as their forecasts for the next realisation.

Even though the mode is the most frequently chosen reply, our results show a substantial variation in the reporting rules used by the individuals. This finding highlights the potential issues regarding the interpretation of forecasts reported in surveys as the respondent's subjective expectations. We demonstrate that these forecasts do not always constitute an accurate measure of the individual's subjective expectations.

To improve the current state of affairs, there are several possible solutions one may implement. Asking for the entire distribution of probabilities may be a possibility in certain contexts. Another idea would be to formulate more explicit questions that would help the survey respondents to understand the answers that are expected from them.

References

[1] Marcel Das, Jeff Dominitz, and Arthur van Soest. Comparing predictions and outcomes: Theory and application to income changes. *Journal of the American Statistical Association*, 94:75–85, 1999.

[2] Adeline Delavande. Probabilistic expectations in developing countries. *Annu. Rev. Econ.*, 6(1):1–20, 2014.

[3] Graham Elliott, Ivana Komunjer, and Allan Timmermann. Biases in macroeconomic forecasts: irrationality or asymmetric loss? *Journal of the European Economic Association*, 6(1):122–157, 2008.

[4] Joseph Engelberg, Charles Manski, and Jared Williams. Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business and Economic Statistics*, 165:146–158, 2009.

[5] Urs Fischbacher. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178, June 2007.

[6] Charles F. Manski. *Identification Problems in the Social Sciences*. Harvard University Press, Harvard, USA, 1999.

[7] Charles F. Manski. Identification of decision rules in experiments on simple games of proposal and response. *European Economic Review*, 46:880–891, 2002.

[8] Charles F. Manski. Measuring expectations. *Econometrica*, 72(5):1329–1376, 2004.

[9] M. Hashem Pesaran and Martin Weale. *Handbook of Economic Forecasting*, volume 1, chapter 14: Survey Expectations, pages 715–776. 2006.

A Instructions

Page 1

Welcome!

Please listen carefully to the instructions. The experiment lasts approximately 120 minutes. During the experiment, we ask that you do not communicate with your neighbours. If you have any questions, please raise your hand and we will answer your question in private.

Page 2

Before starting the experiment, we will present the instructions. We will explain the progress of the experiment in detail.

The experiment consists of two parts that vary slightly in the tasks you will be asked to fulfil. First, we will present the task of the first part and after completing the first part, we will present the task of the second part.

Once you have started, an on-screen summary of the instructions will be available for the duration of the experiment.

We will also provide you with a printed copy of the instructions.

When you have finished the experiment, please, stay seated! We will come to your cubicle and will also give you a reward for your participation.

Page 3

Imagine that you are working for a consulting firm as a forecaster. Your task is to predict the future profits of different companies.

Page 4

All companies make a profit between -35 and 35. More specifically, the profit of each company is one of the following numbers: -35, -25, -5, -5, 5, 15, or 25, 35.

To the left, you can see an example of the chances associated with making these eight profits in the form of a histogram. The vertical axis of this histogram shows the probability that each of the eight potential profits (-35, -25, -15, -5, 5, 15, 25 or 35) realizes.

This histogram is from Company X.

Page 5

You follow the company for 10 periods. In each period, the company makes a new profit. The next period's profit is not known in advance. It will be determined by the computer drawing randomly one from eight profits (-35, -25, -15, -5, 5, 15, 25 or 35) and this is according to the chances that each particular profit is realised. The probabilities that a particular profit is realised are shown in the histogram. The profit of the company in a particular period is independent of other past and future profits.

Your task is to provide a quantified forecast of the company's profit earned in the next period.

As a forecast you can specify any number between -35 and 35.

You can see below an example, where you can enter your prediction for period 4.

Once you have provided your forecasts, you can validate them by pressing the "Validate" button.

Page 6

Once your prediction has been validated, you move on and the profit of the next period is realised as follows:

The computer draws a number at random while respecting the probabilities visible in the histogram. For example, the histogram on this screen indicates that the number "-5" has a chance of ...% being drawn. That is, in a large sample of random draws (e.g. 1,000,000 draws) about ...profits would be "-5".

The computer will draw a new number at each period. The chance that a number will be drawn again corresponds to the percentages indicated in the histogram.

Page 7

The "History" table at the right of the screen summarises for each period the realisations of the profits and your predictions. The "Period" column indicates the period in which the profit was realised. Your prediction for this period is in the "Predictions" column.

The profit made in that period appears in the "Realisation" column after each period.

Page 8

Test the program and make a prediction for period 4 !

Page 9

Note that your prediction appears in the "Predictions" row where period is equal to 4.

In total, you make predictions for 10 companies. This completes the first part. In the second part, you also make forecasts, under other conditions that will be explained to you after completing the first part.

Page 10

A summary of the instructions will be available throughout the experiment. You can consult it by pressing the “Instructions” button at the bottom left. In addition, if necessary, you can access a calculator throughout the experiment by pressing the button with the symbol of a calculator, next to the instructions button.

Page 11

At the end of the experiment, you will receive a summary for all companies that you have evaluated. This information summarises your predictions and profit realisations separately for both parts. You will receive for your predictions 35 CAD and 5 CAD for your participation in the experience - a total of 40 CAD.

Page 12

This is the end of the instructions. If you have any questions, please raise your hand. We will answer your question in private at your place. If you do not have any questions, you can start the experiment.

B Probability Distributions

The distributions used for the experiments $\{p_n\}$ have a discrete support with probability mass on eight equidistant points $n \in \{1, \dots, 8\}$ with distance d. We have chosen a symmetric support around zero with $d = 10$, resulting in the set of points $n \in \{-35, -25, -15, -5, 5, 15, 25, 35\}$ as support for the distributions. The distributions vary in the location of their central tendencies, i.e., the mean $\mu = \sum_{n=\underline{n}}^{\overline{n}} n p_n$, median $m = \{\min n : \sum_{i=\underline{n}}^{n} p_i \geq 0.50\}$ and mode $M = \{n : \max p_n\}$. The distributions are chosen in such a way that their central tendencies lie each in a different interval around a point $n \pm d/2$.

The distributions used in the experiment are presented in Table 3 and Figure 2. In its first three column, the table summarizes the three central tendencies and presents the probability distribution and cumulative distribution. We use each distribution twice: once in its original form (odd lines in Table 3 and left column in Figure 2) and once its mirror image (the following even line and the right column, resp.). For example, distribution 2 is the mirror image of 1. We use two unimodal distributions (no 1 and 3 and their corresponding mirror images 2 and 4).[7] For both distributions $M < m < \mu$, and $\mu < m < M$ for the corresponding mirror image. To allow for the order $m < \mu < M$ (and its mirror image $M < \mu < m$) and to increase the distance between the central tendencies, we additionally use distributions that are "slightly" bimodal. For those distributions, we define as mode the point n with the most probability mass which for distributions 11 and 12 (5 and 6) [7,8 and 9,10] has at least 6 (8) [10] percentage points more than the point with the second highest probability mass. We use distribution 13 as an example in the instructions, but not in the experiment.

In total, counting each distribution and its mirror image separately, we use 12 distributions for the experiment. For half of the distributions, the order of the central tendencies is $M < m < \mu$ (distributions 1, 3 and 5) and $\mu < m < M$ for their mirror distribution. For the other half and their mirror distribution, the order is $M < \mu < m$ (distributions 7, 9 and 11) and $m < \mu < M$, respectively.[8]

[7]A discrete distribution is unimodal with the integer M as mode if $p_n \geq p_{n-1}, n \leq M$ and $p_n \leq p_{n-1}, n > M$. Distributions 1 and 3 (2 and 4, resp.) are not strong unimodal distributions, in the sense of case $p_n^2 \geq p_{n-1}p_{n+1}$. (Encyclopedia of Statistical Sciences, 2006)

[8]Nevertheless, the order $\mu < M < m$ and $m < M < \mu$ is theoretically possible for discrete distributions, we would have needed a support with more equidistant points to implement such order. We did not extend the support in order to have meaningful probability mass on each point of the support.

No interval/bin realization	Central tendencies			Probability and cumulative distribution							
	μ mean	m median	M mode	$-35 \leq x < -30$ -35	$-30 \leq x < -20$ -25	$-20 \leq x < -10$ -15	$-10 \leq x < 0$ -5	$0 \leq x < 10$ 5	$10 \leq x < 20$ 15	$20 \leq x < 30$ 25	$30 \leq x \leq 35$ 35
1	-6.675	-15	-35	0.23 0.23	0.17 0.40	0.1125 0.5125	0.0975 0.61	0.0975 0.7075	0.0975 0.805	0.0975 0.9025	0.0975 1
2	6.675	15	35	0.0975 0.0975	0.0975 0.195	0.0975 0.2925	0.0975 0.39	0.0975 0.4875	0.1125 0.6	0.17 0.77	0.23 1
3	-4.925	-15	-25	0.1125 0.1125	0.23 0.3425	0.17 0.5125	0.0975 0.61	0.0975 0.7075	0.0975 0.805	0.0975 0.9025	0.0975 1
4	4.925	15	25	0.0975 0.0975	0.0975 0.195	0.0975 0.2925	0.0975 0.39	0.0975 0.4875	0.17 0.6575	0.23 0.8875	0.1125 1
5	3	-5	-25	0.01 0.01	0.27 0.28	0.16 0.44	0.08 0.52	0.02 0.54	0.07 0.61	0.19 0.8	0.2 1
6	-3	5	25	0.2 0.2	0.19 0.39	0.07 0.46	0.02 0.48	0.08 0.56	0.16 0.72	0.27 0.99	0.01 1
7	-1.9	15	-35	0.33 0.33	0.1 0.43	0.02 0.45	0.02 0.47	0.02 0.49	0.18 0.67	0.18 0.85	0.15 1
8	1.9	-15	35	0.15 0.15	0.18 0.33	0.18 0.51	0.02 0.53	0.02 0.55	0.02 0.57	0.1 0.67	0.33 1
9	-2.15	5	-25	0.145 0.145	0.205 0.35	0.07 0.42	0.07 0.49	0.145 0.635	0.145 0.78	0.115 0.895	0.105 1
10	2.15	-5	25	0.105 0.105	0.115 0.22	0.145 0.365	0.145 0.51	0.07 0.58	0.07 0.65	0.205 0.855	0.145 1
11	1.8	-15	-25	0.01 0.01	0.3 0.31	0.2 0.51	0.02 0.53	0.03 0.56	0.04 0.6	0.2 0.8	0.2 1

Table 3: Central tendencies (mean, median and mode), probability and cumulative distribution of distributions used for the experiment (1-10) and the distribution used as illustration in the instructions (11).

Discussion Papers of the Research Area Markets and Choice 2019

Research Unit: **Market Behavior**

Azar Abizada, Inácio Bó Hiring from a pool of workers	SP II 2019-201
Philipp Albert, Dorothea Kübler, Juliana Silva-Goncalves Peer effects of ambition	SP II 2019-202
Yves Breitmoser, Sebastian Schweighofer-Kodritsch Obviousness around the clock	SP II 2019-203
Tobias König, Sebastian Schweighofer-Kodritsch, Georg Weizsäcker Beliefs as a means of self-control? Evidence from a dynamic student survey	SP II 2019-204
Rustamdjan Hakimov, Dorothea Kübler Experiments on matching markets: A survey	SP II 2019-205
Puja Bhattacharya , Jeevant Rampal Contests within and between groups	SP II 2019-206
Kirby Nielsen, Puja Bhattacharya, John H. Kagel, Arjun Sengupta Teams promise but do not deliver	SP II 2019-207
Julien Grenet, Yinghua He, Dorothea Kübler Decentralizing centralized matching markets: Implications from early offers in university admissions	SP II 2019-208
Joerg Oechssler, Andreas Reischmann, Andis Sofianos The conditional contribution mechanism for repeated public goods – the general case	SP II 2019-209
Rustamdjan Hakimov, C.-Philipp Heller, Dorothea Kübler, Morimitsu Kurino How to avoid black markets for appointments with online booking systems	SP II 2019-210
Thibaud Pierrot Negotiation under the curse of knowledge	SP II 2019-211
Sabine Kröger, Thibaud Pierrot What point of a distribution summarizes point predictions?	SP II 2019-212

Research Unit: **Economics of Change**

Kai Barron, Steffen Huck, Philippe Jehiel Everyday econometricians: Selection neglect and overoptimism when learning from others	SP II 2019-301

All discussion papers are downloadable:
http://www.wzb.eu/en/publications/discussion-papers/markets-and-choice

Research Unit: **Ethics and Behavioral Economics**

All discussion papers are downloadable:
http://www.wzb.eu/en/publications/discussion-papers/markets-and-choice