Inflation Expectations: The Effect of Question Ordering on Forecast Inconsistencies

Maxime Phillot, Rina Rosenblatt-Wisch

SNB Working Papers
11/2018

SCHWEIZERISCHE NATIONALBANK
BANQUE NATIONALE SUISSE
BANCA NAZIONALE SVIZZERA
BANCA NAZIUNALA SVIZRA
SWISS NATIONAL BANK

Legal Issues

DISCLAIMER

The views expressed in this paper are those of the author(s) and do not necessarily represent those of the Swiss National Bank. Working Papers describe research in progress. Their aim is to elicit comments and to further debate.

LIMITATION OF LIABILITY

The SNB accepts no responsibility for any information it provides. Under no circumstances will it accept any liability for losses or damage which may result from the use of such information. This limitation of liability applies, in particular, to the topicality, accuracy, validity and availability of the information.

ISSN 1660-7716 (printed version)
ISSN 1660-7724 (online version)

Inflation Expectations: The Effect of Question Ordering on Forecast Inconsistencies[a]

Maxime Phillot[b] Rina Rosenblatt-Wisch[c]

July 4, 2018

Abstract

Expectations are key in modern macroeconomics. However, due to their scant measurability, policymakers often rely on survey data. It is thus of critical importance to know the limits of survey data use. We look at inflation expectations as measured through the Deloitte CFO Survey Switzerland and respondents' sensitivity to question ordering thereof. In particular, we investigate whether forecast inconsistencies—the discrepancies between point forecasts and measures of central tendency derived from density forecasts—change significantly depending on whether the point forecast or the density forecast is asked first. We find that a) forecast inconsistencies are sizeable in the data and b) question ordering matters. Specifically, both parametric and non-parametric evaluations of consistency show that c) point forecasts tend to be significantly higher than density forecasts only for those respondents who give a density forecast first. In addition, d) characteristics such as uncertainty, firm size and economic sector relate to inconsistencies.

JEL-Classification: E31, E37, E58

Keywords: Question effects, question ordering, inflation expectations, consistency of forecasts, point forecast, density forecast

[a]We are grateful to Britta Classen, Michael Grampp and Dennis Brandes from Deloitte AG Switzerland for supporting our experiment by implementing it in the Deloitte CFO Survey Switzerland and providing us with the data. We would also like to thank participants of the SNB Brownbag seminar and one anonymous referee for their insightful comments. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Swiss National Bank.

[b]Swiss National Bank, `maxime.phillot@snb.ch`.

[c]Swiss National Bank, `rina.rosenblatt@snb.ch`.

1 Introduction

Expectations are key variables in macroeconomics. However, they are hardly measurable. One way to gauge expectations of either households, professional forecasters or firms is in the form of surveys. At least since Lucas (1972), economists have been widely assuming rational expectations of agents regarding future macroeconomic variables such as income and inflation. In other words, expectations are generally thought to be objectively and optimally formed given all available information. This also means that agents are assumed to be able at all times and under any circumstances to formulate clear and consistent answers. However, cognitive science has documented that seemingly innocuous factors such as purpose of the surveyor, topics covered, ordinary conversational norms, question length, wording and ordering, and many others can have a significant impact on survey responses. These effects are known as question effects and have been studied in various fields, including economics.

To the best of our knowledge, this paper is the first study to address question ordering in inflation expectations. We analyse whether question ordering is crucial for forecast inconsistencies, i.e., the discrepancies between point forecasts and measures of central tendency derived from density forecasts, in inflation expectations. To do so, we make use of the Deloitte CFO Survey Switzerland, which contains two questions about expected inflation in two years' time: the first question asks for a point forecast, the second question asks for a density forecast. From 2014 Q4 until 2017 Q3 we set up an experiment and randomly assigned the order of these two questions to each of the survey respondents. We first assess whether there exists a persistent discrepancy—a forecast inconsistency—between point forecasts and measures of central tendency derived from density forecasts. We then study whether these forecast inconsistencies change significantly depending on the specific order in which these two questions are asked, i.e., point forecast first and density forecasts second or vice versa. We further investigate the relationship between consistency and firm characteristics, such as the size of the firm or the economic sector.

We find that a) forecast inconsistencies are sizeable in the data: approximately 18 to 25 percent of all forecasts are inconsistent. We also find that b) question ordering matters. Asking the density forecast before the point forecast results in an approximately 5 percentage point increase in inconsistencies on average, whereby the question ordering impacts mainly the answers to the point forecast, while the answers to the density forecast seem to be almost unaffected. In addition, c) forecasts are not equally distributed below and above their thresholds of consistency: central tendency measures derived from density forecasts generally reflect lower inflation expectations than point forecasts. This difference is statistically significant mostly for those who are asked the density forecast first. Furthermore, d) characteristics such as uncertainty, firm size and economic sector seem to play a role too: bigger firms from the service sector tend to be more consistent, and higher uncertainty is associated with more inconsistencies.

The remainder of this paper is structured as follows. Section 2 gives an overview of the related literature. Section 3 describes the data and our experiment. Section 4 investigates the effects of question ordering on forecast inconsistencies using non-parametric and parametric methods and shows the results. Section 5 discusses these results and further investigates the relationship between consistency and firm characteristics. Section 6 concludes.

2 Literature Review

2.1 Survey Methodology

That seemingly innocuous factors might affect survey responses is not a new idea. There exists extensive literature in the field of cognitive science, in particular, that addresses respondents' sensitivity to survey methodology. Generally, to the extent that an interview is a form of discussion, it is believed that the principles governing *question effects* should be comparable to those governing the conduct of everyday life conversations (Bradburn, Sudman, and Wansink, 2004).[1] Interviewees are therefore thought to answer to both, the actual questions and the *context* in which they are asked. This includes the purpose of the surveyor, the topics covered by the survey, ordinary conversational norms, etc. However, it also refers to various characteristics such as question length, wording and ordering.[2]

In surveys, we are often interested in respondents' attitudes towards both a general object and specific aspects thereof. This usually translates into asking a general and a specific question consecutively. Consequently, Gobo and Mauceri (2014) argue that this produces an *assimilation effect* causing the respondent to take the first question as the premise of the second. A bias may occur in this context because the respondent aims to answer in a supposedly consistent way. However, the assimilation effect operates in different ways depending on the order of the questions, their degree of connection, and whether they refer to social norms. In this respect, it is believed that when a general and specific question are asked consecutively, answers to the general one are influenced by its position, while those to the specific one are not (Bradburn, Sudman, and Wansink, 2004).[3] Despite the robustness of the latter observation, determining the direction of the bias is not an easy task. In fact, according to Gobo and Mauceri (2014), several sub-phenomena are at play in this context. In short, the assimilation is assumed to be either *inclusive*, in which case the general question summarizes the attitude towards the object of interest (that is, including what the specific question is about), or *exclusive*, in which case the general question refers to everything except the content already implied by the specific question.

Another effect is at play and relates to social norms and values. *Even-handedness* describes the fact that when two items share a common underlying value but one is more popular than the other, asking first about the more popular one strengthens the value so that it holds true for the subsequent item. See for example Rugg and Cantril (1944); Link (1946); Hyman and Sheatsley (1950); Schuman and Ludwig (1983).

[1]For a review on the cognitive psychological theory behind surveys, see Sudman, Bradburn, and Schwarz (1996) or Tourangeau et al. (2000). For a general discussion on attitude measurements, see Petty and Cacioppo (1996), Schwarz (2008), or Petty and Krosnick (2014).

[2]On the theoretical side, Gobo and Mauceri (2014) report that the so-called *cooperative principle* introduced by Grice et al. (1975) underpins question effects. This principle states that participants in a conversation typically presuppose a mutual contribution to it that is accurate, informative and relevant—later stages thus being always tied to what comes before. On the empirical side, potential biases introduced by the context and meaning of adjacent questions have been acknowledged in early works of survey research. Schuman and Presser (1981) provide useful insights into many empirical studies and experiments on question effects.

[3]Why this is the case, on the other hand, remains a "mystery" (Bradburn and Danis, 1984, p. 117).

2.2 Economic Surveys

In economics, surveys of opinion constitute a typical instrument to measure expectations of market participants, policymakers, and researchers. In the light of our previous discussion, there is reason to believe that survey methodology also matters for survey responses in economic surveys. Moreover, there exists an empirical body in the economic literature that points towards the presence of question effects. For instance, Bruine de Bruin et al. (2012) study the effect of question wording regarding inflation expectations of households.[4] They run an experiment where they randomly assign respondents one of three different ways of wording inflation: "prices in general", "prices you pay" and "inflation". Comparing the respective medians through Mann-Whitney-Wilcoxon tests, they find that "prices you pay" and "prices in general" yield similar responses, but of higher expectation and disagreement than questions about "inflation". The authors argue that respondents are more prone to think about extreme personal price experiences when asked about "prices in general" and "prices you pay" than about "inflation", the latter performing better in pinning down the theoretical concept.

Another empirical body of research studies forecast biases and inconsistencies by comparing point forecasts with density forecasts, questions which, for example, are included in the Survey of Professional Forecasters (SPF). To collect this type of forecast, the questionnaire invites interviewees to assign probabilities to several suggested intervals of future realizations of inflation. Conceptually speaking, answering such questions amounts to drawing a histogram, providing the surveyor with a discrete distribution over expected inflation. Although no parameter underlying the distribution is explicitly reported by the respondent, it is possible to infer some of them—such as the mean and variance—in various ways depending on the assumptions about the distribution one makes.[5]

The early literature comparing point and density forecasts explored *uncertainty*.[6] Zarnowitz and Lambros (1987) first addressed the soundness of a then common practice of associating *consensus*—the dispersion of point predictions—with *uncertainty*—the diffuseness of the corresponding probability distributions. The goal was to investigate whether a high degree of consensus among forecasters reflects general certainty, or in other words, if the *interpersonal* dispersion is a good proxy for the *intrapersonal* one. They found that the standard deviation of point forecasts tends to understate the mean dispersion of individual density forecasts, although they remain generally positively correlated. Giordani and Söderlind (2003) followed by comparing and discussing the relevance of both measures plus a third one, the variance of aggregate histograms, in capturing uncertainty.[7] They argue that they are all relevant depending on what one wishes to capture and find that disagreement (i.e., lack of consensus) is a reasonable proxy for uncertainty.

However, another concern that has been raised more recently in the literature comes from the observation that point forecasts and measures of central tendency derived from density forecasts do not always match. These apparent *forecast inconsistencies* have been acknowledged first by Engelberg, Manski, and Williams (2009), who assessed consistency using both a

[4]Initial results can be found in Van der Klaauw et al. (2008) and Bruine de Bruin et al. (2010).

[5]For a survey on density forecasts, their applications, evaluations and limits, see Tay and Wallis (2000).

[6]A comprehensive list of economic research using SPF data is available on Philadelphia Fed's website.

[7]Aggregate histograms are obtained by averaging over individuals the probability assigned to each bin.

non-parametric and a parametric approach. The non-parametric approach consists of verifying solely if the point forecast lies within a subjective range whose bounds are constructed using the lower and upper bounds of the questionnaire, respectively.[8] Accordingly, consistency is either true or false. In that sense, the non-parametric approach yields an agnostic assessment of consistency but gives no information as to the *degree* of consistency a particular individual shows. In contrast, the parametric techniques explicitly extract parameters from density forecasts, but they require assumptions on the underlying distribution. Looking at non-parametric subjective means, medians and modes, Engelberg et al. (2009) observe that approximately 10 to 20 percent of the responses are inconsistent both regarding GDP growth and inflation, but less so for latter quarters of the year. They argue that this is due to the nature of the SPF, which asks about end-of-year forecasts—the horizon of the forecasts being thus reduced every quarter.

More interestingly, they find that among those point forecasts that are inconsistent with their respective density forecast, a higher proportion underestimates inflation, and overestimates GDP growth. This tendency to submit an overly optimistic point forecast is confirmed by the parametric approach which, by fitting generalized beta distributions to the density forecasts, shows that approximately 40 percent of point predictions are higher than their fitted median for GDP growth, whereas approximately 70 percent are below regarding inflation.

Other contributions point towards the fact that professional forecasters are not necessarily internally consistent and tend to provide point forecasts that are rosier than their density forecast, see, e.g., Clements (2009), Garcia and Manzanares (2007) or Boero, Smith, and Wallis (2008). Overall, the results are in line with those of Engelberg et al. (2009) and Clements (2009) in that reporting biases are quite frequent and generally reflect an overly optimistic scenario, i.e., insufficient inflation and excessive GDP growth.

What could rationalize forecasters' inclination towards favourable outcomes? A hypothesis that has been put forward is that forecasters use asymmetric loss functions. The idea that forecasters may face different costs whether they over- or underestimate future variables originates as a response to the critique that forecasts are recurrently biased, suggesting the existence of a bound on rationality.[9] Clements (2010) rather advocates in favour of bounded rationality.

Another hypothesis could be that neither the mean, median nor mode is reported by the forecaster. After all, none of the aforementioned surveys asks for specific statistics, but rather for a best prediction, which can be heterogeneously interpreted among respondents as a formal statistic or an informal appraisal. Meyler and Rubene (2009) and Stark (2013) study the effect of judgement and show that forecasters rely on both mathematical models and judgement to form their forecasts.[10]

[8] See section 4.1 for more details.

[9] See Conlisk (1996) for the critique, and notably Elliott, Komunjer, and Timmermann (2008), and Capistrán and Timmermann (2009) for the asymmetric loss hypothesis.

[10] The ECB and the Philadelphia Fed respectively issued a special questionnaire to gauge how their panellists compute and provide their predictions. The ECB reports that interviewees on average weight *judgement* as contributing up to 40 percent of their forecast. Approximately 80 percent of the respondents produce their density forecast solely based on judgement. When asked about which statistic they refer to for their point forecast, approximately 75 percent checked the mean, 20 percent the median, and 7 percent the mode. The

All in all, these findings taken together point towards economic surveys being subject to question effects. In this paper, we focus on the sensitivity of survey responses to question ordering in regard to inflation expectations. In particular, our goal is to investigate whether question ordering helps in understanding the amount of forecast inconsistencies in surveys about future inflation. In other words, do the discrepancies between point forecasts and measures of central tendency derived from density forecasts change significantly depending on whether the point forecast or the density forecast is asked first?

3 Data

3.1 The Deloitte CFO Survey

In this paper, we use data from the Deloitte CFO Survey conducted in Switzerland at a quarterly frequency since the third quarter of 2009. The survey covers the views of Chief Financial Officers (CFOs) and Group Financial Directors of companies in Switzerland from all relevant sectors on their outlook for business, as well as on financing, risks and strategies. The number of respondents varies but is usually over a hundred firms. Participants are emailed each quarter to fill in the questionnaire. However, for reasons of anonymity, Deloitte does not provide the individual identifiers. We are thus unable to exploit the panel structure. In addition, the panel of participating CFOs changes as well.

The survey is conducted online, and the participants do not have to provide answers for all the questions to complete the survey. The survey covers 20 questions that recur each quarter and approximately 10 questions unique to the financial conditions of the previous quarter.[11]

Since our focus lies on inflation expectations, we will mainly look at the two following questions:

1. *In two years' time, what annual rate of inflation, as measured by the Swiss consumer price index, do you expect?*
2. *In two years' time, where do you expect the annual rate of inflation (Swiss Consumer Price Index) to be?*

$$(-\infty,-4],\ (-4,-2],\ (-2,-1],\ (-1,0],\ (0,1],\ (1,2],\ (2,4],\ (4,+\infty)$$

The first question asks for a point estimate of two-year-ahead annual inflation rate (in percent), while the second offers a fixed number of intervals for the same rate, to which respondents are requested to assign probabilities. These intervals together form a symmetric eight-bin centred histogram. At the interval level, we interpret missing values as zeros. If the assigned probabilities do not add up to 100% we normalize them so that they add up to 100%. Moreover, for our analysis we exclude observations where either answer is missing. Appendix A.1 provides additional information about missing observations, the assigned probabilities and their normalization.

In addition to inflation expectations, the questionnaire provides information on the responding firm. In particular, three questions allow us to know more about the size, the openness and the sector of the firm:

Philadelphia Fed presents a similar picture: 80 percent of their interviewed panellists revealed that they rely on both mathematical models and judgement to form their forecasts.

[11] For more details see `www2.deloitte.com`.

3. *What was your company's turnover in the last financial year?*
4. *How much of your company's revenues are earned outside Switzerland?*
5. *In which sector does your company primarily operate?*

Question 3 offers several intervals that we group into two categories: less than CHF 500 million (*low* turnover), and CHF 500 million or more (*high* turnover). Question 4 answers are regrouped as follows: less than one third (*low* share), and one third or more (*high* share). Question 5 suggests a list of several "sectors" from which respondents are allowed to select more than one answer. We group all combinations into three sectors: construction, manufacturing, and services.[12]

3.2 The Experiment

As of 2014 Q1 we implemented the following experiment together with Deloitte: until then, questions 1 and 2 were always asked in the same order—point forecast first, density forecast second. From 2014 Q1 the order was assigned randomly to participants. The implementation took some time. Until 2014 Q3 we had to assign the ordering manually as follows. First, participants had to answer the point forecast, then the density forecast. We then switched the ordering after approximately 50% of the CFOs whom we expected to participate in the respective quarter concluded the survey. We are fully aware that these two groups might have had quite different information sets each quarter. This in turn could have influenced their answers on inflation expectations. We therefore treat 2014 Q1 until 2014 Q3 as a trial period. From 2014 Q4 onwards, the computer program was adjusted such that the order of the two questions was completely randomized with no manual interference, giving each respondent a true 50% chance of seeing the point before the density forecast or the other way around. On the computer screen, the participants only see one question at a time. The following analysis of inconsistency of the forecasts will be based on the sample with complete randomization, i.e., from 2014 Q4 until 2017 Q3.

Table 1 shows the summary statistics of the data we analyse. It shows the number of observations that were first assigned the point, respectively the density forecast and their respective sample mean. It also gives an overview of the average assigned probability for each bin of the density forecast. In addition, it reports details regarding the turnover, openness and sector of the firms. The statistical analysis of the differences between the group that was first asked a density forecast and the group that was first asked a point forecast and of their forecast inconsistencies is the subject of the following section.

4 Estimation & Results

4.1 Methodology

Generally, a forecaster is said to be internally consistent if he or she gives the same answer to two identical questions asked differently. In our case, each point forecast can reasonably be thought to match some statistic derived from the respective subjective probability distribution function underlying expectations over future inflation, which in turn should be summarized

[12] The groups are constructed to match the statistical classification of economic activities in the European Community (NACE) at best.

Table 1: Deloitte CFO Survey Summary Statistics

Variable	Observations			Mean			Std. dev.		
	N	N_D	N_P	μ	μ_D	μ_P	σ	σ_D	σ_P
Inflation Expectations									
1) Point Forecast	**1268**	**631**	**637**	0.84	0.92	0.76	0.65	0.63	0.67
Two-years-ahead inflation expectation									
2) Density Forecast	**1268**	**631**	**637**						
Probability that two-years-ahead inflation lie within...									
$(-\infty, -4]$				0.09	0.11	0.07	0.96	1.29	0.46
$(-4, -2]$				0.54	0.45	0.64	2.25	1.59	2.75
$(-2, -1]$				3.33	3.46	3.19	8.15	8.67	7.60
$(-1, 0]$				16.83	17.49	16.19	17.33	18.28	16.33
$(0, 1]$				47.41	47.27	47.54	24.71	24.95	24.49
$(1, 2]$				25.09	24.37	25.82	20.98	20.82	21.12
$(2, 4]$				6.04	6.14	5.93	9.61	10.02	9.18
$(4, +\infty)$				0.66	0.71	0.62	3.73	4.46	2.85
Attributes									
3) Turnover	**1255**	**625**	**630**						
For the last financial year (millions CHF)									
$[0, 50)$	240	129	111						
$(50, 100]$	191	86	105						
$(100, 500]$	378	180	198						
$(500, 1000]$	148	74	74						
$(1000, +\infty)$	298	156	142						
4) Openness	**1222**	**616**	**606**						
Share of revenues earned abroad									
$[0, 1/3)$	517	256	261						
$[1/3, 2/3)$	118	56	62						
$[2/3, 1]$	587	304	283						
5) Sector	**1256**	**624**	**632**						
Construction	103	56	47						
Manufacturing	590	294	296						
Services	563	274	289						

Notes: Each cell represents the number of observations (N_i), the sample mean (μ_i), or the sample standard deviation σ_i from group $i = P, D$ for the sample between 2014 Q4 and 2017 Q3, i.e., during the experiment. P (D) denotes the respondents who were asked for a point (density) forecast first.

in each density forecast. In other words, if we knew each respondent's forecasting model and the statistic reported as the point forecast, we should be able to map density forecasts into point forecasts almost exactly. We could then confidently consider any difference as a *forecast inconsistency*. Unfortunately, with the data at hand we need to make assumptions to match density forecasts with their respective point forecasts.

There are two approaches to assessing consistency between point forecasts and density forecasts: the non-parametric and the parametric one. The non-parametric approach binds consistency by using the edges of each interval given in the survey but makes no further assumption regarding the underlying subjective distribution. The parametric approach, however, explicitly states the shape of the distribution and may rely on fitting techniques to obtain its parameters such as the mean and variance. The fundamental difference between these two methods lies in whether one wishes to assume how the probability mass is distributed *within* each bin. We therefore face a trade-off: While the non-parametric approach provides a more agnostic assessment, it does not give any information as to the *degree* of inconsistency one forecaster might show. In particular, under the non-parametric approach, we are only able to say whether a forecast is consistent or not, whereas the parametric approach tells us exactly by how much.

As for the parametric approach, we will follow Zarnowitz and Lambros (1987). This widely applied approach only assumes that the probability mass of density forecasts is located at the centre of each bin. This allows us to compute the *midpoint* of each density forecast, i.e., its subjective mean.[13] A technical requirement however is to close the interval of the first and the last bin. In Question 2 (see Section 3) the first and the last bin is formulated as a one-sided open interval. To close the interval, we attribute the value -6 and 6, as it reproduces the length of 2 percentage points of inflation of the intervals, respectively, following and preceding them.[14] A drawback of this methodology is that it over-evaluates the variance under bell-shaped densities. In this respect, the so-called Sheppard's correction may help to obtain a more realistic estimate of the variance but is only computable if the bins are of the same size, which is not the case in the Deloitte CFO Survey. Notwithstanding, because we are not required to accurately evaluate the uncertainty surrounding density forecasts in our set-up, we chose to follow the abovementioned approach for its readability and simplicity.[15]

The following example should illustrate the difference between both approaches: If a forecaster assigns the probabilities 0.3, 0.4, 0.2, 0.1 to the bins $(-1, 0]$, $(0, 1]$, $(1, 2]$, $(2, 4]$ respectively and 0 elsewhere, then the non-parametric approach binds midpoint consistency between $-1 \cdot 0.3 + 0 \cdot 0.4 + 1 \cdot 0.2 + 2 \cdot 0.1 = 0.1$ and $0 \cdot 0.3 + 1 \cdot 0.4 + 2 \cdot 0.2 + 4 \cdot 0.1 = 1.2$. The forecast is then considered consistent if the point forecast lies within $(0.1, 1.2]$. The lower (upper) bound accounts for the possibility that the forecaster always considered the lowest (highest) value of the bin while reporting the probabilities. By contrast, the parametric approach infers that the subjective midpoint be exactly $-0.5 \cdot 0.3 + 0.5 \cdot 0.4 + 1.5 \cdot 0.2 + 3 \cdot 0.1 = 0.65$, because it supposes that the forecaster always and exclusively considered the centre of the bin. Any deviation of the point forecast from this value can then be associated with inconsistency.

As the point estimate question does not specify what statistic of the subjective probability distribution the respondent should report, forecasters might report the median of their subjective distribution as their point forecast rather than the midpoint. To account for this case, we computed subjective medians as follows. In the non-parametric case, the subjective median is the first interval itself whose cumulative probability is 50 percent or more. In the parametric

[13] Assuming the mass is uniformly distributed within each bin produces equivalent midpoint estimates.

[14] This choice is virtually irrelevant, since only 2.5 percent of the treatment sample assigned a probability greater than or equal to 10% to either of the extreme bins. All our results remain robust for other choices.

[15] Appendix A.2 describes an alternative approach which consists in fitting normal distributions to individual density forecasts by numerical optimization as in Giordani and Söderlind (2003). All our results are robust to such methodology as shown in Appendix A.3.

Table 2: Midpoint and Median Forecast Consistency by Question Ordering

	Subjective Midpoint						**Subjective Median**					
	Consistent		Below		Above		Consistent		Below		Above	
Quarter	λ_P^c	λ_D^c	λ_P^b	λ_D^b	λ_P^a	λ_D^a	λ_P^c	λ_D^c	λ_P^b	λ_D^b	λ_P^a	λ_D^a
2014 Q4	80.3	82.1	9.8	1.8	9.8	16.1	72.1	71.4	14.8	1.8	13.1	26.8
2015 Q1	81.4	66.7	8.5	10.5	10.2	22.8	74.6	73.7	16.9	12.3	8.5	14.0
2015 Q2	74.5	62.0	14.5	8.0	10.9	30.0	67.3	60.0	21.8	4.0	10.9	36.0
2015 Q3	83.7	73.1	4.1	3.8	12.2	23.1	79.6	73.1	8.2	5.8	12.2	21.2
2015 Q4	77.2	71.9	17.5	1.8	5.3	26.3	78.9	73.7	15.8	1.8	5.3	24.6
2016 Q1	87.7	80.4	8.8	5.9	3.5	13.7	77.2	78.4	12.3	7.8	10.5	13.7
2016 Q2	76.5	81.5	17.6	3.7	5.9	14.8	52.9	64.8	29.4	7.4	17.6	27.8
2016 Q3	86.0	71.2	6.0	5.8	8.0	23.1	84.0	59.6	8.0	13.5	8.0	26.9
2016 Q4	80.4	81.3	7.8	4.2	11.8	14.6	80.4	79.2	5.9	4.2	13.7	16.7
2017 Q1	80.4	76.4	5.9	0.0	13.7	23.6	76.5	72.7	9.8	1.8	13.7	25.5
2017 Q2	75.6	79.6	8.9	2.0	15.6	18.4	68.9	75.5	17.8	6.1	13.3	18.4
2017 Q3	96.1	78.0	2.0	0.0	2.0	22.0	80.4	70.0	9.8	2.0	9.8	28.0
Pooled	81.6	75.3	9.4	4.0	8.9	20.8	74.4	71.0	14.3	5.7	11.3	23.3

Notes: Each cell represents the percentage λ_i^k of respondents from group $i = P, D$ falling in the category $k = c, b, a$, for the quarter in row. The subscripts P (D) denotes the respondents who were asked for a point (density) forecast first. The superscripts c, b, a respectively denote whether the point forecast lies within, below or above its level of consistency.

case, it is the middle of the same interval.[16]

4.2 Non-Parametric Approach

Table 2 displays the results of the non-parametric approach. For each quarter of the experiment and by question ordering, it shows the percentage of respondents that gave a point forecast respectively within, below or above their respective interval of consistency, evaluated according to the above described non-parametric subjective midpoints and according to the non-parametric subjective medians. We denote such percentages by λ_i^k, where $i = P, D$ stands respectively for the group of respondents who were asked for a point forecast or a density forecast first, and $k = c, b, a$ stands respectively for consistent, below and above. The last row depicts the pooled sample.

Focusing on midpoints, we observe a proportion of consistency that ranges from 74.5 to 96.1 percent for the P group, and from 62 to 82.1 for the D group. Quarterly consistency between the groups correlates by 19.4 percent, which indicates that time-varying macro factors exert a common pressure on consistency although in a relatively low manner. Looking at the pooled sample tells us that those respondents who reported their point forecast before their density forecast were consistent in 81.6 percent of the cases, whereas those who reported their density forecast first were consistent in 75.3 percent of the cases. In other words, consistency (as defined by the non-parametric approach) of the P group exceeded that of the D group, on average, by a 6.3 percentage points margin. This difference of 6.3 percentage points is statistically significant at the 1% level as seen in Table 10 in the Appendix A.4.

More interestingly, a quick inspection of inconsistent forecasts reveals that the proportion of point forecasts that lie above and below their respective level of consistency is quite het-

[16]Equivalently, one might be interested in assessing mode consistency. Because this requires further assumptions, we detail such analysis and show the robustness of our results thereto in Appendix A.3, Table 7.

erogeneous and depends on question ordering. For those who were asked for a point forecast first, the amount of under-evaluations of point forecasts relative to density forecasts ranges between 2 and 17.6 percent, while this amount ranges only between 0 and 10.5 percent for the other group. Conversely, the proportion of over-evaluated point forecasts varies from 2 to 15.6 percent for the P group, while it goes from 14.6 to as much as 30 percent for the D group.

In total, those who reported a point forecast first understated inflation slightly more often (9.4 percent below versus 8.9 percent above), while the others almost systematically overstated inflation (4 percent below versus 20.8 percent above). In other words, being asked for the density forecast before the point forecast not only increases the amount of inconsistency but also makes it more likely for the point forecast to overstate the level of inflation as suggested by the density forecast.

The scrutiny of median non-parametric consistency gives the same general message. Interestingly, consistency occurs less often in the data when we evaluate consistency based on the relationship between point forecasts and subjective medians. This may indicate that forecasters actually link their point forecast to the mean of their density forecast rather than to the median thereof. Notwithstanding, the P group remains more consistent than the D group by 3.4 percentage points. However, this difference is not significant as we show in Table 10 in the Appendix A.4. Moreover, the pattern in the discrepancies between excessively high and excessively low point forecasts as a function of question ordering is preserved.

Overall, these results provide evidence that question ordering matters. In particular, asking for a point forecast before a density forecast seems to result in fewer occurrences of inconsistency. Furthermore, it appears that asking first for a point (density) forecast produces a slight (strong) tendency to report point forecasts reflecting a lower (higher) level of inflation than the respective subjective midpoints and medians. Therefore, our non-parametric assessment of consistency indicates that the effect of question ordering is twofold, for it both strongly impacts the *amount* of inconsistencies and their *nature*.

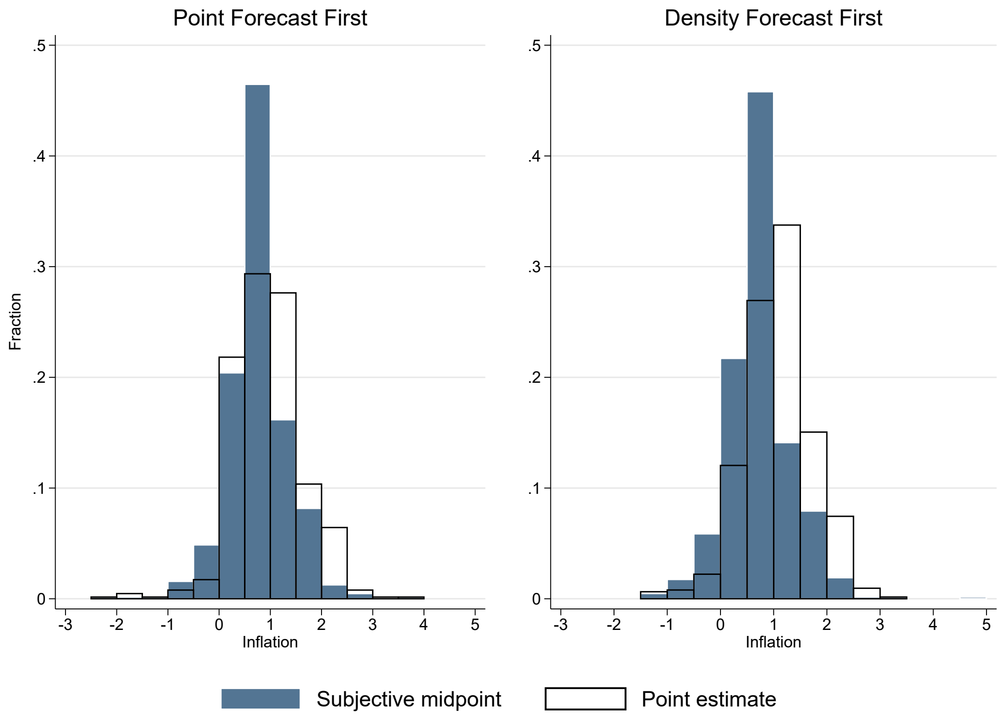
4.3 Parametric Approach

The parametric approach allows us to derive from the density forecasts some measures of central tendency that are in levels. However, as detailed above, it requires assumptions. The measure we are focusing on in our analysis is the midpoint, i.e., the subjective mean of density forecasts under the assumption that the probability mass is exactly located at the centre of each bin.

Figure 1 plots, for each question ordering, the histogram of subjective midpoints against the histogram of point forecasts in the pooled sample (from 2014 Q4 to 2017 Q3). In particular, it shows the fraction of respondents who reported a forecast corresponding to a certain level of inflation (in bins of size 0.5 percentage points), either directly (subjective point forecasts, in translucent-white) or indirectly (subjective midpoints, in blue).

On the one hand, it appears that the distribution of subjective midpoints is quite homogeneous between the groups (i.e., comparing the left and the right panel), with a fraction of almost 70 percent of all midpoints being comprised between 0 and 1 percent of inflation. On the other hand, however, it seems that the distribution of point forecasts shifts towards the

FIGURE 1: Point Forecasts and Subjective Midpoints by Question Ordering



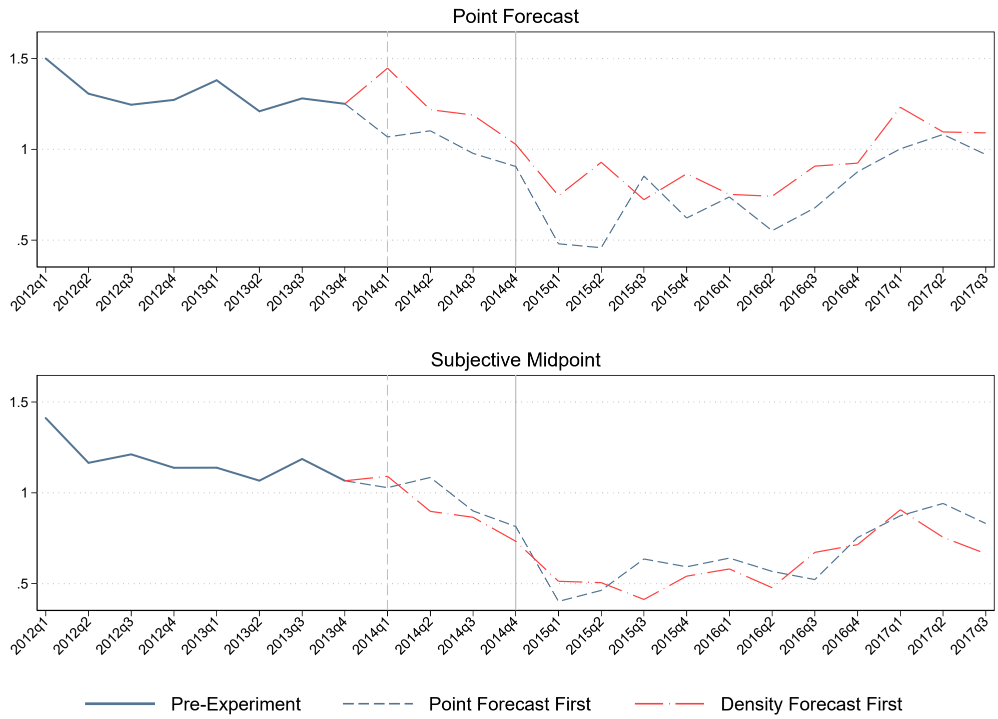
Notes: The figure plots, for each group, the fraction of respondents (from 2014 Q4 to 2017 Q3) who reported a forecast corresponding to a certain level of inflation (in bins of size 0.5 percentage points), either directly (point forecast, in white) or indirectly (subjective midpoint derived from density forecast, in blue).

centre of the distribution of subjective midpoints when one jumps from the right to the left panel. Indeed, while approximately 50 percent of point forecasts lie between 0 and 1 percent of inflation for the P group, only approximately 35 percent do for the D group.

Clearly, forecasters being asked for the point forecast before the density forecast generally give a point forecast that is more in line with the density forecast than forecasters facing the opposite ordering. In other words, we can already confirm the result from the non-parametric analysis, that forecasters tend to be less consistent when they first report a density forecast.

Figure 2 breaks down point forecasts and subjective midpoints by group and quarterly averages, and plots them as a time series. The dashed vertical line marks the implementation of the experiment (trial period), while the solid vertical line marks the starting point of our analysis (i.e., from 2014 Q4 to 2017 Q3). The blue solid line represents quarterly averages of point forecasts, respectively subjective midpoints for the pre-experiment period. The blue dashed line depicts quarterly averages of point forecasts, respectively subjective midpoints for the P group and the red dashed dotted line the ones for the D group. Recall that before the implementation of the experiment (i.e., from 2012 Q1 to 2014 Q1), point forecasts were always asked before density forecasts. The blue dashed line, which shows the results of the group that sees the point forecast question first (P group), can therefore be expected to follow the pattern of the solid blue line (like a control group) – any deviation of the red dashed dotted

Figure 2: Quarterly Point Forecasts and Subjective Midpoints



Notes: The dashed vertical line marks the implementation of the experiment, while the solid vertical line marks the starting point of our analysis.

line from the blue dashed line can thus be interpreted as the effect of flipping the question ordering (i.e., the treatment effect). Comparing the point forecasts between the two groups shows that the average point forecast of the D group is somewhat persistently higher than that of the P group (Figure 2, top panel). For the average midpoint one can barely distinguish the two series (Figure 2, bottom panel).

Table 3 formalizes these observations for the pooled series. First, it shows the sample mean point forecast of the D and the P group and the respective standard deviation. The 0.16 percentage points difference in the mean point forecasts between the two groups is statistically significant, while the 0.94 ratio between their respective standard deviations is not.

Second, Table 3 shows the mean probabilities assigned to each bin for both groups. The mean probabilities assigned to each bin can never be said to differ significantly between the groups. However, we reject equal variance of the assigned probability between groups for all but two of the eight bins. Interestingly, these two bins together comprise inflation from above zero to below two percent, and account on average for more than 70 percent of cumulated probability. Given that the Swiss National Bank defines price stability as an annual inflation rate below 2% but positive, this result suggests that credibility by forecasters about the capacity of the central bank to achieve its target is not affected by question ordering. In other words, inflation expectations seem to be too well anchored regarding "normal territories" for question ordering to affect the variability of its associated probability (which, as we noted

Table 3: Comparison Between Groups

Variable	Mean			Std. dev.		
	μ_D	μ_P	$\mu_D - \mu_P$	σ_D	σ_P	σ_D/σ_P
Inflation Expectations						
1) Point Forecast *Two-years-ahead inflation expectation*	0.92	0.76	0.16*** (4.3)	0.63	0.67	0.94 (0.9)
2) Density Forecast *Probability that two-years-ahead inflation lie within...*						
$(-\infty, -4]$	0.11	0.07	0.04 (0.8)	1.29	0.46	2.8*** (8.0)
$(-4, -2]$	0.45	0.64	−0.19 (−1.5)	1.59	2.75	0.58*** (0.3)
$(-2, -1]$	3.46	3.19	0.27 (0.6)	8.67	7.60	1.14*** (1.3)
$(-1, 0]$	17.49	16.19	1.3 (1.3)	18.28	16.33	1.12** (1.2)
$(0, 1]$	47.27	47.54	−0.27 (−0.2)	24.95	24.49	1.02 (1.0)
$(1, 2]$	24.37	25.82	−1.45 (−1.2)	20.82	21.12	0.99 (0.9)
$(2, 4]$	6.14	5.93	0.21 (0.4)	10.02	9.18	1.09* (1.2)
$(4, +\infty)$	0.71	0.62	0.09 (0.4)	4.46	2.85	1.56*** (2.5)
Subjective midpoint	0.66	0.68	−0.02 (−0.6)	0.61	0.60	1.03 (1.1)

Notes: This Table complements Table 1. It shows the sample mean (μ_i) and the sample standard deviation σ_i from group $i = P, D$ for the sample between 2014 Q4 and 2017 Q3, i.e., during the experiment. P (D) denotes the respondents who were asked for a point (density) forecast first. t and F statistics respectively for the mean- and variance-comparison tests are given in parentheses. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

earlier, is centred around the same value for both groups). What question ordering does seem to affect is the (lack of) consensus as to the probability of rarer events, i.e., disinflation and high inflation. Nonetheless, identifying a pattern is difficult, for the significant differences in standard deviations between the two groups only reflect a cumulated probability of 30 percent.

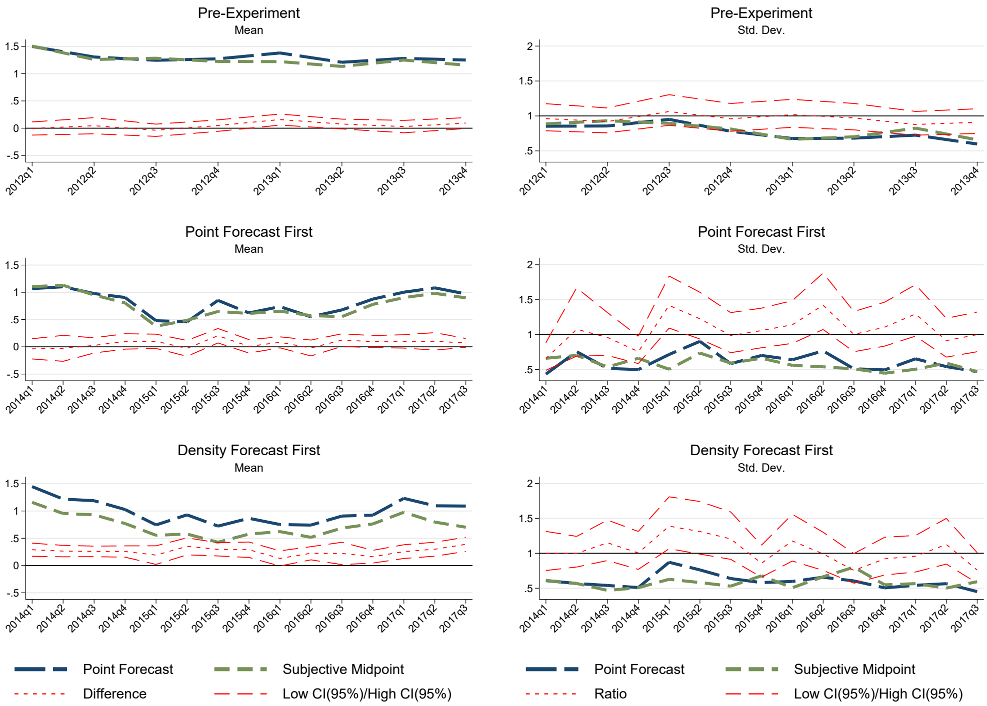
Third, Table 3 reports the sample mean subjective midpoints of the density forecasts for both groups. The −0.02 percentage points difference in the sample mean midpoint between the two groups is not statistically significant.

Overall, the results suggest that the point forecast question reacts much more sensitively to question ordering than the density forecast. Asking for a density seems to be a more robust way of retrieving inflation expectations as it seems to be unaffected by question ordering.

In a next step, we analyse the differences between point forecasts and subjective midpoints within each group, i.e., the so-called forecast inconsistencies. The panels on the left of Figure 3 plot as a time series the quarterly averages of point forecasts (blue dashed lines) and subjective midpoints (green dashed lines), as well as the differences between the two (i.e., the forecast inconsistencies, red dotted lines) respectively for the pre-experiment sample (top panels) and the experiment sample broken down by question ordering (middle and bottom panels).[17] The red dashed lines surrounding the series of mean forecast inconsistencies are the lower and upper bounds of the 95 percent confidence interval (CI) for the difference between the mean of point

[17]Note that Figure 3 also shows the trial period of the experiment (2014 Q1 – 2014 Q3) for each group although, as we argued before, it does not provide a reliable assessment of the treatment effect. All the comments exposed here therefore do *not* consider this period, despite the robustness in doing so.

Figure 3: The Effect of Question Ordering on Forecast Inconsistencies



Notes: Left panels plot quarterly averages of point forecasts and subjective midpoints, as well as their differences together with the 95% CI bands thereof assuming equal variances. Right panels plot quarterly standard deviations of the same variables, as well as their ratio together with the 95% CI bands thereof. Each row considers a different subsample: pre-experiment (2012 Q1 – 2013 Q4), and experiment (2014 Q1 – 2017 Q3) by question ordering.

forecasts and the mean of subjective midpoints, computed separately for each quarter through two-sample mean-comparison t-tests assuming equal variances. In a very similar fashion, the right panels show the quarterly standard deviations of point forecasts and subjective midpoints as well as their ratios. The dashed lines surrounding these ratios are the bounds of the 95 percent CI thereof, computed separately for each quarter through two-sample variance-comparison F-tests.

Focusing first on the left panels of Figure 3 allows us to assess the effect of question ordering on forecast consistency. A quick comparison between the top and middle panels tells us that the P group indeed follows the pattern of the pre-experiment sample. In fact, similar to prior to the experiment, those who submitted a point forecast first during the experiment provided on average point forecasts sometimes higher, sometimes lower than their respective midpoints, but for a difference that can almost never be considered significantly different from zero.

By contrast, forecasters from the D group systematically submitted point forecasts that were higher on average than their subjective midpoints. Strikingly, this overstatement of inflation made by point forecasts relative to density forecasts is statistically significant at the quarterly level for almost every period. We thus observe a strong treatment effect: switching the order by asking for the density forecast before the point forecast exerts an upward

Table 4: Forecast Inconsistencies and Treatment Effect

	Obs.		**Inconsistency**		**Treatment Effect**	
Quarter	N_D	N_P	Δ_D	Δ_P	$\Delta_D - \Delta_P$	p-value
2014 Q4	56	61	0.26	0.10	0.16	0.04
2015 Q1	57	59	0.19	0.10	0.09	0.20
2015 Q2	50	55	0.35	-0.03	0.38	0.00
2015 Q3	52	49	0.30	0.20	0.10	0.15
2015 Q4	57	57	0.28	0.00	0.28	0.00
2016 Q1	51	57	0.13	0.08	0.05	0.29
2016 Q2	54	51	0.22	-0.02	0.24	0.00
2016 Q3	52	50	0.22	0.12	0.10	0.21
2016 Q4	48	51	0.16	0.10	0.06	0.21
2017 Q1	55	51	0.25	0.10	0.15	0.04
2017 Q2	49	45	0.30	0.10	0.20	0.03
2017 Q3	50	51	0.39	0.07	0.32	0.00
Pooled	631	637	0.26	0.08	0.18	0.00

Notes: The Table displays, for each quarter of the experiment, the number of respondents N_i and the average inconsistency Δ_i for each group $i = P, D$ as well as the difference thereof. The last column displays the p-value of the t-test that this difference $\Delta_D - \Delta_P$ is positive, under the null hypothesis that it is zero (assuming equal variances). The last row considers the pooled sample.

pressure on point forecasts relative to midpoints, thereby producing an increase in forecast inconsistencies.

Nevertheless, to give a formal appraisal of the average treatment effect and its significance, we need to go one step further and compare the average inconsistencies *between* the two orderings. This is comparable to a difference in differences approach: because point forecasts are on average higher than subjective midpoints for both groups as shown in Figure 3, only the difference between the respective discrepancy captures the causal effect of question ordering. To this end, Table 4 summarizes by quarter the number of respondents N_i and the average forecast inconsistency Δ_i for each group $i = P, D$ as well as the difference thereof, which captures the average treatment effect. The last column displays the p-value of the t-test that this difference $\Delta_D - \Delta_P$ is positive, under the null hypothesis that it is zero (assuming equal variances).

For every quarter of the experiment, the average treatment effect is positive. In seven out of twelve quarters, it is significantly so at the 95 percent level. The pooled sample tells us that the discrepancy between point forecasts and subjective midpoints is on average positive and significantly higher by 0.18 percentage points of inflation for the D group than for the P group. Clearly, imposing an alternative ordering by requesting a density forecast before a point forecast causes forecast inconsistencies to widen significantly.

Finally, looking at the right panels of Figure 3 provides an indication of the plausibility of our results. The standard deviation of point forecasts (or subjective midpoints) is a measure of *disagreement* and is often used in the literature as a proxy for general uncertainty.[18] We interpret the quasi-permanent conservation of the null hypothesis (i.e., that standard deviations

[18]Because the quarterly sample size of each question ordering is half the size of the pre-experiment sample, the volatility of the series becomes mechanically lower.

are equal) for both question orderings as evidence that question ordering affects the amount of inconsistencies, but not the general level of disagreement among forecasters. In other words, asking for a density forecast first intensifies the discrepancies between point forecasts and midpoints, but does so without distorting their respective dispersion. We can thus exclude that the experiment itself came as a surprise which would in turn drive our results.

All in all, the results from the parametric approach confirm those of the non-parametric one by unambiguously pointing towards the presence of question effects in surveys about inflation expectations. In particular, we find that asking for the density forecast before the point forecast results almost systematically in a statistically significant discrepancy between point forecasts and midpoints, with point forecasts overstating the level of inflation suggested by the density forecast. By contrast, asking for the point forecast first appears to produce differences between midpoints and point forecasts that are of no statistical significance. The difference in the level of consistency between the two question orderings ultimately underscores the causal interpretation of our results.

5 Discussion

5.1 Is Anchoring at Play?

In the previous section, we observed that asking for the density forecast before the point forecast results almost systematically in a statistically significant discrepancy between point forecasts and subjective midpoints, with point forecasts overstating the level of inflation suggested by density forecasts. Could some form of anchoring be at play? One may argue that for a respondent who first sees the density, the point estimate will likely be anchored to the range that was shown in the density question. If anchoring is at play, we would expect the following for those, who give a density forecast first: i) a lower variance of the point forecasts and ii) a distribution of the point forecasts centred around zero due to the symmetry of the bins. It appears that the latter hypothesis can be discarded by looking at Table 3. The mean point forecasts of the D group are further away from zero than those of the P group. The former hypothesis seems to apply to a certain extent. The point estimates from respondents who give a density forecast first have an overall standard deviation of 0.04 lower (in terms of inflation) than the other group. However, this difference is not statistically significant.[19] As far as the data can tell, anchoring did not cause the point estimates to be more narrowly distributed. Of course, to study the effects of anchoring thoroughly, we would have to run further treatments. However, this goes beyond the scope of this paper and the data at hand do not allow us to draw further conclusions.

5.2 Regression Analysis

What drives forecast inconsistencies? As we have already noted, question ordering does. However, other factors such as firm's characteristics or uncertainty might very well be influencing the discrepancy between density forecasts and point forecasts. To address this question, we make use of the firm's attributes present in our data, define a measure of uncertainty and estimate two models: a logistic regression and a linear regression.

[19] Actually, the null hypothesis of equal variance cannot be rejected in any single quarter.

Table 5: Recoding Attributes in Binary Variables

Variable	Observations		
	N	N_D	N_P
Attributes			
3) Turnover	**1255**	**625**	**630**
For the last financial year (millions CHF)			
$0 = [0, 500)$	809	395	414
$1 = [500, +\infty)$	446	230	216
4) Openness	**1222**	**616**	**606**
Share of revenues earned abroad			
$1 = [0, 1/3)$	517	256	261
$0 = [1/3, 1]$	705	360	345
5) Sector	**1256**	**624**	**632**
0 = Construction & Manufacturing	693	350	343
1 = Services	563	274	289

Notes: The table displays the number of observations for each attribute after recoding them in binary variable. For the original data, see Table 1.

Recall that we have information about a) the turnover, b) the share of revenues earned abroad and c) the operating sector of the respondent's firm. To be parsimonious, we recode these three attributes into binary variables. For each of them, Table 5 displays the threshold we chose as well as the number of observations falling in each category by question ordering. Note, that neither group is over- or underrepresented in terms of their question ordering, so that we can exclude that attrition is correlated with the attributes.[20]

We coded the dummy variables so that we expect the value 1 to be associated with more consistency. First, we consider a turnover greater than or equal to CHF 500 million to be a high turnover. A higher turnover should reflect a higher size and access to better data, or a greater need for quality forecasts. Second, we define a share of revenues earned abroad between zero and one third as low openness. Arguably, a domestically oriented firm is more likely to depend on national rather than international prospects, and thus to monitor local prices accurately. Finally, firms from the services sector could be associated with higher levels of technology or financial market knowledge, and thus with more rigorous forecasts.

As a measure of uncertainty at the individual level, we argue as Clements (2010) that the number of bins that are assigned a positive probability by the respondent is a good proxy for the variance of the density function underlying forecasters' expectations over future inflation. The advantage of this measure is that it is non-parametric and readily available.[21] Similarly, we recode this variable as a dummy whose value is 1 if the number of bins used by the respondent is lower than or equal to 3 (i.e if the forecast is of high certainty) and 0 otherwise.

We turn now to our baseline model, the logistic regression (logit). In this respect, suppose we have N independent realizations $\{y_j\}_{j=1,...,N}$ of a random variable Y_j. Let $Y_j \sim \text{Bernoulli}(\lambda_j^c)$ and y_j be equal to 1 if respondent j is consistent and 0 otherwise. We can

[20] Table 11 in the Appendix A.5 explores the correlations between firm attributes. There is little correlation present in the data.

[21] Appendix A.3 explores the robustness of our results to using parametric evaluations of subjective dispersion.

then model the probability λ_j^c using a linear predictor function according to

$$\text{logit}(\lambda_j^c) = d_j \alpha + x_j' \beta + z_j \gamma, \tag{1}$$

where d_j is a dummy for the treatment group, x_j a vector of attributes, z_j a measure of uncertainty, and α, β, γ a set of parameters. We then estimate the regression coefficients in Equation (1) through maximum likelihood estimation.

This specification makes use of our non-parametric assessment of consistency. The idea here is to predict the likelihood that a forecaster will produce a consistent forecast based on his or her question ordering, the characteristics of his or her employing firm, and the uncertainty surrounding his or her forecast. However, since the coefficients that Equation (1) yields are log odds and thereby difficult to interpret, we compute and report the marginal probability changes evaluated at means associated with a discrete change away from the reference category. Because we coded our binary regressors such that switching away from the reference category (i.e., from zero to one) should increase the probability of being consistent, our estimates will tell us by how much it does at the margin for an average respondent.[22]

As an alternative model, we use our parametric assessment of consistency and estimate the following linear regression (LR):

$$-|\Delta_j| = d_j \tilde{\alpha} + x_j' \tilde{\beta} + z_j \tilde{\gamma} + \tilde{\varepsilon}_j, \tag{2}$$

where $|\Delta_j|$ is the absolute difference between the point forecast and the midpoint given by respondent j, $\tilde{\alpha}$, $\tilde{\beta}$, $\tilde{\gamma}$ a new set of parameters, and $\tilde{\varepsilon}_j$ are i.i.d. errors.

Considering the distance in absolute terms and negating it makes our two specifications comparable, because it recovers our notion of consistency in levels. In particular, all else equal, each coefficient can be interpreted as the average marginal increase in closeness between subjective midpoints and point forecasts produced by switching away from the reference category of the dummy regressors. While Equation (1) provides estimates to be interpreted in terms of probabilities, Equation (2) yields estimates in terms of percentage points of inflation. Therefore, the linear regression model will serve us both as an assessment of the robustness of the results under the logit, and as an indication of their economic significance.

Table 6 displays the results from our two specifications based either on midpoint consistency (columns (1) and (2)) or on median consistency (columns (3) and (4)). Note that to account for potential global time-varying factors, we include in all our models time-fixed effects.[23,24] As mentioned above, the logit coefficients (odd columns) show the marginal increase in the probability of being consistent induced by a discrete change of the variable in row, when all the other variables take their mean value. LR coefficients (even columns) express the average percentage point increase in closeness between the centre of the density forecast and the point forecast associated with the same change.

[22] Recall that the logistic transform is a nonlinear combination of the regressors, so that we need to fix their value to assess marginal probability changes. We take their sample respective means.

[23] A caveat of our approach, however, is that we cannot control for individual fixed effects. This is not a problem insofar as forecasters' ability to be consistent through time is *not* correlated with our regressors. In other words, we need to make the assumption that this ability is unobservable by employers and that there is no self-selection of better forecasters into certain types of firms. Since this may be argued to be a somewhat strong assumption, one should interpret our estimates as upper bounds.

[24] All our results are robust to the non-inclusion of time-fixed effects.

Table 6: Logit and LR of Midpoint and Median Consistency on Attributes

	Subjective Midpoint		**Subjective Median**	
	Logit	LR	Logit	LR
	(1)	(2)	(3)	(4)
Point Forecast First	0.0635**	0.0716**	0.0285	0.102**
	(2.82)	(3.96)	(1.06)	(4.36)
High Certainty	0.0723**	0.0513	0.103***	0.0726**
	(3.21)	(1.84)	(5.54)	(3.24)
High Turnover	0.0668***	0.0560***	0.0333	0.0439**
	(3.69)	(4.61)	(1.27)	(3.34)
Low Openness	0.00959	0.00818	0.00964	0.0438*
	(0.40)	(0.37)	(0.55)	(2.62)
Services Sector	0.0352*	0.0416	-0.0194	-0.00636
	(2.06)	(1.54)	(-1.00)	(-0.36)
Constant		-0.502***		-0.597***
		(-19.95)		(-22.63)

Notes: t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. $N = 1217$. Logistic regression (logit) models use the proportion of consistent forecasts as the dependent variable, whereas linear regression (LR) models use the negative absolute difference between the point forecast and the central tendency measure derived from the density forecast. Logit coefficients represent marginal probability changes evaluated at means. All models include time fixed effects, and standard errors are clustered at the quarterly level.

Overall, the results confirm our previous findings that question ordering matters. Focusing on subjective midpoints first, column (1) indicates that an average forecaster (in term of its other characteristics) is as much as 6.35 percent more likely to submit a consistent forecast if he or she is asked for a point forecast first. Column (2) tells us that such ordering makes the point forecast on average closer to the subjective midpoint by 7.16 basis points.[25]

In addition, Table 6 suggests that consistency depends on some firms' attributes and certainty as well. First, being more certain induces a marginal increase of 7.23 percent in the probability of being consistent. However, it does not seem to exert a significant effect on the closeness between subjective midpoints and point forecasts. Second, if the respondent works in a firm with a high turnover, the probability marginally increases by 6.68 percent, and reduces the distance between the midpoint and the point forecast by 5.6 basis points. Third, we cannot say that openness has an impact on consistency in either specification. Fourth, although consistency in levels does not seem to significantly vary with the sector in which the firm operates (column (2)), it appears that switching to the service industry marginally increases the probability for the average forecaster to be consistent by 3.52 percentage points. Finally, the constant reflects part of our previous results, saying that a rather uncertain forecaster working in a small open construction or manufacturing firm hands in a point forecast on average 0.5 percentage points away from the midpoint if he or she submits a density forecast first.

We now inspect median consistency to assess the robustness of these results. The logistic regression in this context (column (3)) globally suggest a similar picture but with somewhat

[25] A basis point is a hundredth of a percent of inflation.

less statistical significance. In fact, only certainty turns out to significantly raise the probability of consistency. Moreover, the service sector dummy now exerts a negative marginal effect—although not significant—on such probability. Column (4) on the other hand reveals that the linear regression model performs better than its counterpart from column (2). We can indeed infer that all our dummy variables except for the sectoral one provokes a positive and significant effect on forecast consistency as measured by the closeness between subjective point forecasts and the median of the density forecasts.

Interestingly, the positive effect on consistency of question ordering and certainty is of higher magnitude than in column (2). This result together with the non-significance of the corresponding coefficient in column (3) indicates that question ordering makes little difference in the marginal probability that the misalignment between the median and the point forecast exceeds a relevant threshold but stills makes this misalignment on average greater by 10.2 basis points when the density forecast is asked first. In addition, the constant term reveals a greater discrepancy than in column (2). This reinforces our previous argument that subjective midpoints capture the information relevant to point forecasts better than medians.

5.3 Interpretation

We unambiguously showed one should consider question effects in economic surveys. In particular, we find strong evidence that question ordering distorts the internal consistency of two-year-ahead inflation forecasts. However, question ordering not only affects the *amount* of inconsistencies, it also influences the *direction* in which the mismatch occurs. Moreover, the results from our regression analysis indicate that, beyond question ordering, several factors can affect the amount and the level of inconsistencies. Specifically, we find that more certainty and greater firm size (as measured by the turnover) are associated with more consistencies in a robust manner. Some evidence suggests that sector and openness of firms also affect consistency, but with somewhat less significance.

Are our results in line with the literature on question effects we laid out in Section 2? Note, all surveys being analysed so far had the same ordering: point forecast first, density forecast second. In line with this literature we find that forecast inconsistencies persistently occur. The literature on forecast inconsistencies also finds that point forecasts tend to underestimate inflation with respect to their density forecast. Our findings regarding underestimation for the ordering point forecast first, density forecasts second are mixed. Our results of the non-parametric approach tend slightly towards underestimation, while our results of the parametric approach tend slightly towards overestimation.

Our contribution to this literature lies in the additional insight question ordering brings to the debate. When we switch the order of the questions, i.e., when the density forecast precedes the point forecast, we detect a clear overestimation both for the non-parametric and for the parametric approach. We observe that mainly the answers to the point forecast were affected by the switch in the order, while the density forecasts remained robust. The literature on question effects in cognitive science finds the following explanation: Answers to a general question are affected by its position, while those to a specific or more detailed one are not. If one interprets the point forecast to be a general question about two-years-ahead inflation and the density forecast to be a specific or more detailed one thereabout, then our results are in line with cognitive science.

To the extent that qualitative investigations about the way forecasts are produced revealed a great reliance on "judgement" and that forecasters are likely to be heterogeneous, our results should not come as a surprise. It is rather unclear which of the mean, median, mode or even quantile of the density forecast should match the point forecast. Our findings suggest that when designing a survey, if one wishes to minimize inconsistencies, one should place the question of the point forecast first. If one is limited in the number of questions that can be included in a survey, one should opt for a question on density forecasts. These sorts of questions seem to be less sensitive to question ordering, and in addition, they reveal more information.

6 Conclusions

We showed that question ordering matters in economic surveys and is relevant for questions on inflation expectations. While the answers to the point forecast were sensitive to the order in the survey, the answers to the density forecast were basically unaffected. We found that inconsistencies between the point forecasts and measures of central tendency derived from density forecasts are sizeable in the data and are increased if respondents submit a density forecast before a point forecast. Firms' characteristics seem to play a role as well in the sense that size and economic sector relate to inconsistencies, as does uncertainty.

These results suggest that the design of surveys also matters in regard to economics. When gauging expectations on macroeconomic variables from surveys, policymakers and market participants alike should be aware that biases due to question effects might be at play. This should not imply that surveys are not a useful policy instrument; on the contrary, they deliver additional information compared to market data, or sometimes cover areas where no market data exist. If one is limited in the number of questions that can be included in a survey, we suggest focusing on density forecasts since they seem to be more robust to question effects, and, by their nature, they convey more information than simple point forecasts.

References

G. Boero, J. Smith, and K. F. Wallis. Uncertainty and disagreement in economic prediction: the Bank of England Survey of External Forecasters. *The Economic Journal*, 118(530): 1107–1127, 2008.

N. Bradburn and C. Danis. Potential contributions of cognitive research to survey questionnaire design. *Cognitive aspects of survey methodology: Building a bridge between disciplines*, pages 101–129, 1984.

N. M. Bradburn, S. Sudman, and B. Wansink. *Asking questions: the definitive guide to questionnaire design–for market research, political polls, and social and health questionnaires.* John Wiley & Sons, 2004.

W. Bruine de Bruin, W. Vanderklaauw, J. S. Downs, B. Fischhoff, G. Topa, and O. Armantier.

Expectations of inflation: The role of demographic variables, expectation formation, and financial literacy. *Journal of Consumer Affairs*, 44(2):381–402, 2010.

W. Bruine de Bruin, W. Van der Klaauw, G. Topa, J. S. Downs, B. Fischhoff, and O. Armantier. The effect of question wording on consumers' reported inflation expectations. *Journal of Economic Psychology*, 33(4):749–757, 2012.

C. Capistrán and A. Timmermann. Disagreement and biases in inflation expectations. *Journal of Money, Credit and Banking*, 41(2-3):365–396, 2009.

M. P. Clements. Internal consistency of survey respondents' forecasts: Evidence based on the Survey of Professional Forecasters. *The methodology and practice of econometrics. A festschrift in honour of David F. Hendry*, pages 206–226, 2009.

M. P. Clements. Explanations of the inconsistencies in survey respondents' forecasts. *European Economic Review*, 54(4):536–549, 2010.

J. Conlisk. Why bounded rationality? *Journal of Economic Literature*, 34(2):669–700, 1996.

G. Elliott, I. Komunjer, and A. Timmermann. Biases in macroeconomic forecasts: irrationality or asymmetric loss? *Journal of the European Economic Association*, 6(1):122–157, 2008.

J. Engelberg, C. F. Manski, and J. Williams. Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business & Economic Statistics*, 27(1):30–41, 2009.

J. A. Garcia and A. Manzanares. Reporting biases and survey results: evidence from European professional forecasters. Technical report, ECB Working Paper No. 836, 2007.

P. Giordani and P. Söderlind. Inflation forecast uncertainty. *European Economic Review*, 47 (6):1037–1059, 2003.

G. Gobo and S. Mauceri. *Constructing survey data: an interactional approach.* Sage, 2014.

H. P. Grice, , et al. *Syntax and Semantics, Volume 3: Speech Acts*, chapter Logic and Conversation, pages 41–58. Academic Press, 1975.

H. H. Hyman and P. B. Sheatsley. The current status of American public opinion. *The teaching of contemporary affairs*, pages 11–34, 1950.

H. C. Link. The Psychological Corporation's index of public opinion. *Journal of Applied Psychology*, 30(4):297, 1946.

R. E. Lucas. Expectations and the neutrality of money. *Journal of Economic Theory*, 4(2): 103–124, 1972.

A. Meyler and I. Rubene. Results of a special questionnaire for participants in the ECB Survey of Professional Forecasters (SPF). Technical report, MPRA Paper No. 20751, 2009.

R. E. Petty and J. T. Cacioppo. *Attitudes and persuasion: Classic and contemporary approaches*. Westview Press, 1996.

R. E. Petty and J. A. Krosnick. *Attitude strength: Antecedents and consequences*. Psychology Press, 2014.

D. Rugg and H. Cantril. The wording of questions. *Gauging public opinion*, pages 23–50, 1944.

H. Schuman and J. Ludwig. The norm of even-handedness in surveys as in life. *American Sociological Review*, pages 112–120, 1983.

H. Schuman and S. Presser. Questions and answers: Experiments on question form, wording, and context in attitude surveys. *New York: Academic*, 1981.

N. Schwarz. Attitude measurement. *Attitudes and attitude change*, 3:41–60, 2008.

T. Stark. SPF panelists' forecasting methods: A note on the aggregate results of a November 2009 special survey. *Federal Reserve Bank of Philadelphia*, 2013.

S. Sudman, N. M. Bradburn, and N. Schwarz. *Thinking about answers: The application of cognitive processes to survey methodology.* Jossey-Bass, 1996.

A. S. Tay and K. F. Wallis. Density forecasting: a survey. *Journal of Forecasting*, 19(4):235, 2000.

R. Tourangeau, L. J. Rips, and K. Rasinski. *The psychology of survey response.* Cambridge University Press, 2000.

W. Van der Klaauw, W. Bruine de Bruin, G. Topa, S. Potter, and M. F. Bryan. Rethinking the measurement of household inflation expectations: preliminary findings. Technical report, Federal Reserve Bank of New York Staff Report No. 359, 2008.

V. Zarnowitz and L. A. Lambros. Consensus and uncertainty in economic prediction. *Journal of Political Economy*, 95(3):591–621, 1987.

A Appendix

A.1 Data

We compile our data by assembling Deloitte's quarterly surveys into a larger dataset. Because we focus on forecast inconsistencies, we drop all observations for which either the point forecast or the density forecast on inflation expectations is missing (the survey does not force the box to be filled in). This occurred 246 times out of 1,514 for the experiment sample, and 129 out of 1,251 for the pre-experiment one. A missing density forecast occurs when none of the intervals is used. When at least one interval contains a positive probability, we interpret unused intervals as zero-probability intervals.

Furthermore, the probabilities assigned to the intervals occasionally do not add up to a 100 percent (the survey does not require answers to do so). For 93.5 percent of all observations however, the probabilities add up to 100 percent. For 97.7 percent of them, their sum is comprised between 90 and 110 or is equal to one. All the remaining observations range between 0.3 and 500. Nevertheless, to conserve the full information of our sample, we normalize all the probabilities so that they add up to 100 percent.

A.2 Distribution Fitting

As mentioned in Section 4, assuming that all the mass of the density forecast lies at the centre of each bin tends to overstate the level of uncertainty if the underlying distribution is thought to be bell-shaped. Although we do not make explicit use of the second moment of the density forecasts, it is worth considering an alternative parametric approach to assess the robustness of our results.

Thus, following Giordani and Söderlind (2003), we can assume that each forecaster's density forecast is normally distributed, and solve for each individual parameter through numerical optimization. Formally, we would like to estimate for each respondent $j \in \{1, ..., N\}$ the subjective mean μ_j and the subjective variance σ_j^2 according to

$$\min_{\hat{\mu}_j, \hat{\sigma}_j^2} \sum_{k=1}^{K} (P[L_k < Z_j \leq U_k] - p_{j,k})^2,$$

where K is the number of bins, $Z_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$, L_k and U_k respectively denote the lower and upper bound of bin k, and $p_{j,k}$ the probability associated by respondent j to bin k. In other words, we pick the set of parameters that minimizes the sum of squared differences between the probability mass lying under the curve of a normal density following these parameters, and the probability mass assigned by the respondent.

When the number of bins used by the respondent does not exceed two, the fitting may not be satisfying. To address this issue in a simple manner we assume 1) a uniform distribution within the bin if only one bin is used, and 2) that the mass lies at the centre of the bin if exactly two bins are used. The first assumption avoids a subjective variance of 0. Note that this procedure slightly differs from the one used by Giordani and Söderlind (2003) since we have to address bins of different sizes. However, these two special cases occur in less than 20 percent of our experiment sample, and thus should not be of critical importance.

Overall, this specification is somewhat more restrictive than the one we use in the paper, as it assumes that the underlying distribution is symmetric and unimodal. Nevertheless, it is appealing in that it equates the mean, the mode and the median. Appendix A.3 shows the results associated with this approach.

A.3 Robustness

Non-Parametric Mode.— One could argue that the mode of the density forecast is reported as the point forecast rather than the midpoint or the median. To address the plausibility of such a hypothesis, we apply the same non-parametric exercise to this statistic.

The non-parametric subjective mode is taken as the bin itself to which is assigned the highest probability. When the highest probability is assigned to more than one bin, we take the bin that is closest to the midpoint. We do so because it prevents the need to address cases involving bins of different sizes, or cases of multi-modal density forecasts.

For each quarter and by question ordering, Table 7 displays the proportion of respondents whose point forecast lies respectively within, below or above its consistency level.

TABLE 7: Mode Consistency by Question Ordering

	Subjective Mode					
	Consistent		Below		Above	
Quarter	λ_P^c	λ_D^c	λ_P^b	λ_D^b	λ_P^a	λ_D^a
2014 Q4	73.8	73.2	16.4	3.6	9.8	23.2
2015 Q1	74.6	75.4	18.6	12.3	6.8	12.3
2015 Q2	70.9	64.0	20.0	4.0	9.1	32.0
2015 Q3	85.7	73.1	10.2	7.7	4.1	19.2
2015 Q4	77.2	71.9	17.5	7.0	5.3	21.1
2016 Q1	80.7	74.5	14.0	13.7	5.3	11.8
2016 Q2	64.7	64.8	29.4	7.4	5.9	27.8
2016 Q3	82.0	59.6	10.0	13.5	8.0	26.9
2016 Q4	86.3	79.2	9.8	6.3	3.9	14.6
2017 Q1	78.4	65.5	13.7	9.1	7.8	25.5
2017 Q2	73.3	79.6	17.8	8.2	8.9	12.2
2017 Q3	80.4	74.0	11.8	2.0	7.8	24.0
Pooled	77.2	71.2	15.9	7.9	6.9	20.9

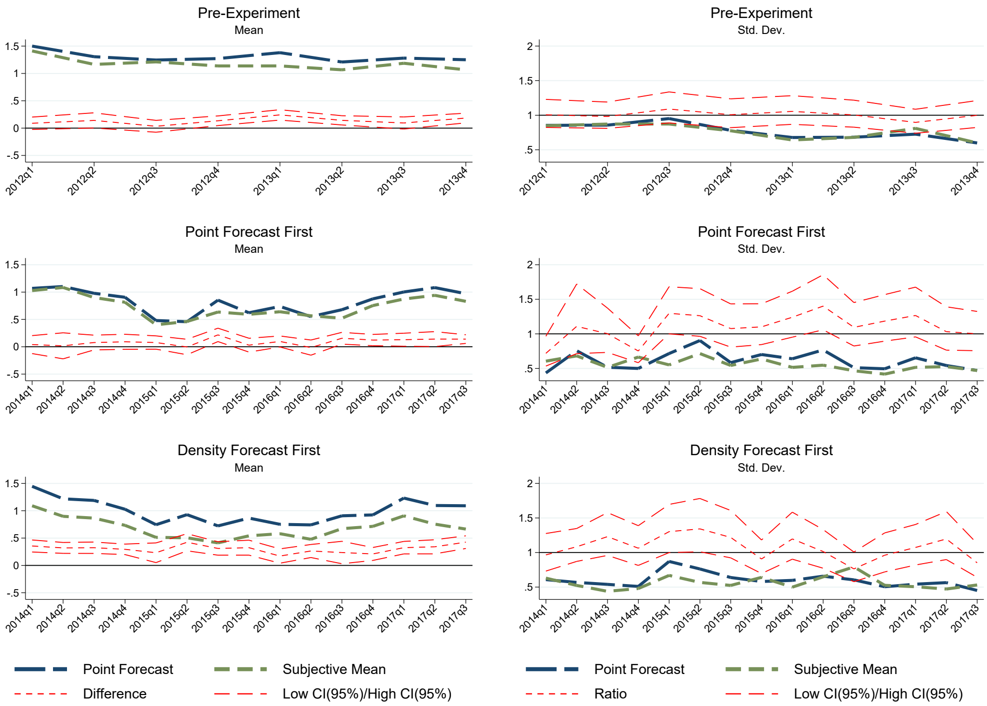
Notes: See Table 2 in the paper for details.

All the conclusions drawn from Table 2 are conserved. Asking for a point forecast first yields point forecasts that are more frequently mode-consistent than asking for a density forecast first by a 6 percentage points average margin. Moreover, we observe a stronger discrepancy between the two question orderings in regard to inconsistent forecasts. In particular, an inconsistent point forecast is way more likely to lie below its consistent level if the point forecast is asked first, but way more likely to lie above it if the density forecast is asked first.

Normally Fitted Parameters.— Figure 4 shows the effect of question ordering on forecast inconsistencies when we use the subjective means stemming from the normal density fitting approach described in Appendix A.2 instead of the subjective midpoints.

Clearly, the picture yields the same general interpretation as to the effect of question ordering on forecast inconsistencies. For the experiment sample, although we observe a slightly higher degree of inconsistencies for the P group compared to using midpoint (see Figure 3 in the paper), these quarterly average inconsistencies remain of rather low statistical significance.

FIGURE 4: Treatment Effect Under An Alternative Parametric Approach



Notes: See Figure 3 in the paper for details.

By contrast, the null hypothesis of equal means between point forecasts and subjective means can still be rejected for every single quarter regarding the D group. The treatment effect under this specification remains qualitatively unchanged, as indicated by Table 8. Indeed, with a significant average difference in inconsistencies of 0.19 percentage points of inflation, we reject the null hypothesis that this difference is zero at the quarterly level in seven out of twelve cases.

Table 9 displays the results from the linear regression estimated in Equation (2) when we use the subjective means from the fitted normal distributions instead of the midpoints in the computation of the dependent variable.

The cells are therefore the coefficients from the linear regression of the absolute difference in negative terms between the point forecasts and the subjective means on question ordering, certainty and firm characteristics. Column ($1'$) considers the exact same variable of certainty as in the paper (c.f. Table 6), while column ($2'$) considers an alternative dummy variable based on the subjective standard deviation derived from the normal fitting approach. Namely, its value is one if the subjective standard deviation is less than or equal to 0.6, and zero otherwise. The value of 0.6 was chosen because it is the median subjective standard deviation in the full sample.

Comparing column ($1'$) here with its counterpart in Table 6 (i.e., column (2)) leads to the exact same conclusions. Furthermore, looking at column ($2'$) reinforces the view that more certainty (at the individual level) is associated with a higher degree of consistency. Using

Table 8: Forecast Inconsistencies and Treatment Effect

	Obs.		**Inconsistency**		**Treatment Effect**	
Quarter	N_D	N_P	Δ_D	Δ_P	$\Delta_D - \Delta_P$	p-value
2014 Q4	56	61	0.31	0.09	0.22	0.01
2015 Q1	57	59	0.24	0.09	0.15	0.09
2015 Q2	50	55	0.41	0.01	0.39	0.00
2015 Q3	52	49	0.31	0.22	0.09	0.15
2015 Q4	57	57	0.32	0.04	0.29	0.00
2016 Q1	51	57	0.17	0.10	0.08	0.18
2016 Q2	54	51	0.25	-0.01	0.26	0.00
2016 Q3	52	50	0.24	0.16	0.08	0.25
2016 Q4	48	51	0.21	0.13	0.08	0.16
2017 Q1	55	51	0.33	0.12	0.20	0.01
2017 Q2	49	45	0.34	0.14	0.19	0.02
2017 Q3	50	51	0.42	0.14	0.28	0.00
Pooled	631	637	0.29	0.10	0.19	0.00

Notes: See Table 4 in the paper for details.

Table 9: LR of Subjective Mean Consistency on Attributes

	Subjective Mean	
	(1′)	(2′)
Point Forecast First	0.0870***	0.0862***
	(4.68)	(4.93)
High Certainty	0.0425	0.105**
	(1.53)	(3.26)
High Turnover	0.0537**	0.0559**
	(3.94)	(4.41)
Low Openness	0.0230	0.0148
	(1.19)	(0.71)
Services Sector	0.0142	0.0107
	(0.47)	(0.36)
Constant	-0.486***	-0.515***
	(-16.75)	(-17.75)

Notes: See Table 6 in the paper for details.

a parametric measure of certainty, namely, the subjective standard deviation recoded into a dummy variable, makes the corresponding coefficient highly significant.

Overall, we argue that all our results are robust to adopting the alternative parametric approach described in Appendix A.2, which consists in fitting normal distributions to individual density forecasts. Using the subjective means instead of the midpoints yields the same general conclusions.

A.4 Treatment Effect in Non-Paramteric Approach

In Table 2 we show, by group, the percentage of respondents whose point forecast is respectively consistent with, below and above the central tendency of the corresponding density forecast. For the sake of clarity, the table does not show the significance of the differences between groups. Doing so would amount to assessing the treatment effect of our experiment, which we document in a more sophisticated way through the parametric approach (namely, in Table 4).

Since it still may be of interest, Table 10 displays the differences between the two groups in the proportions along with the p-value for the corresponding (not shown) t-statistics. The null hypothesis is that the proportion of respondents is the same regardless of the question ordering, and the test assumes equal variance.

Table 10: Group Differences in Forecast Consistency

	Subjective Midpoint						**Subjective Median**					
	Consistent		Below		Above		Consistent		Below		Above	
Quarter	$\Delta\lambda^c$	p	$\Delta\lambda^b$	p	$\Delta\lambda^a$	p	$\Delta\lambda^c$	p	$\Delta\lambda^b$	p	$\Delta\lambda^a$	p
2014 Q4	-1.8	0.80	8.1	0.07	-6.2	0.32	0.7	0.93	13.0	0.01	-13.7	0.06
2015 Q1	14.7	0.07	-2.1	0.71	-12.6	0.07	0.9	0.91	4.7	0.48	-5.6	0.35
2015 Q2	12.5	0.17	6.5	0.30	-19.1	0.01	7.3	0.44	17.8	0.01	-25.1	0.00
2015 Q3	10.6	0.20	0.2	0.95	-10.8	0.16	6.5	0.45	2.4	0.64	-8.9	0.24
2015 Q4	5.3	0.52	15.8	0.00	-21.1	0.00	5.3	0.51	14.0	0.01	-19.3	0.00
2016 Q1	7.3	0.30	2.9	0.57	-10.2	0.06	-1.2	0.88	4.4	0.45	-3.2	0.61
2016 Q2	-5.0	0.53	13.9	0.02	-8.9	0.14	-11.9	0.22	22.0	0.00	-10.1	0.22
2016 Q3	14.8	0.07	0.2	0.96	-15.1	0.04	24.4	0.01	-5.5	0.38	-18.9	0.01
2016 Q4	-0.9	0.91	3.7	0.45	-2.8	0.68	1.2	0.88	1.7	0.70	-2.9	0.69
2017 Q1	4.0	0.62	5.9	0.07	-9.9	0.20	3.7	0.66	8.0	0.08	-11.7	0.13
2017 Q2	-4.0	0.64	6.8	0.14	-2.8	0.72	-6.6	0.48	11.7	0.08	-5.0	0.51
2017 Q3	18.1	0.01	2.0	0.32	-20.0	0.00	10.4	0.23	7.8	0.10	-18.2	0.02
Pooled	6.4	0.01	5.5	0.00	-11.8	0.00	3.4	0.17	8.6	0.00	-12.0	0.00

Notes: $\Delta\lambda^k$ is the difference $(\lambda_P^k - \lambda_D^k)$ between the two groups in the proportions shown in Table 2 for each category $k = c, b, a$, that is, whether the point forecast lies within, below or above its level of consistency. p denotes the p-value for the t-test that this difference is zero, assuming equal variance.

As seen in Table 10, the difference in midpoint-consistency ($\Delta\lambda^c$) is statistically significant when one considers the pooled sample (last row). This observation is also true for the difference in mean inconsistencies both for the share of forecasts lying below and above their consistent level. In other words, respondents who are asked first for a density forecast tend to be less consistent (as opposed to the other group) and when they are inconsistent, they tend to overestimate inflation more and to underestimate it less.

In regard to median consistency, it appears that the difference in the pooled sample ($\Delta\lambda^c$) is not significant, although the differences in the distribution of inconsistencies ($\Delta\lambda^b$ and $\Delta\lambda^a$) are. This generally confirms our previous finding that when a point forecast is inconsistent, it is more likely to be above (below) its consistency level if the density forecast was produced first (second).

As for quarterly differences, they are hardly significant. This can be generally explained by the small sample size. Nevertheless, all the figures, when they are significant, offer a consistent view and lead to the same conclusions as above.

A.5 Correlation Between Attributes

Table 11 displays correlations between firm attributes displayed in Table 5 together with the significance thereof. As shown in the table, there is little correlation present in the data. In fact, only firm sector and openness correlate significantly, unsurprisingly as service firms tend to be more domestically oriented.

Table 11: Correlation Between Attributes

	High Turnover	High Openness	Service Sector	High Certainty
High Turnover	1			
High Openness	0.04	1		
Service Sector	-0.04	-0.30***	1	
High Certainty	0.02	0.00	-0.00	1

Notes: *** $p < 0.001$. Each entry is the pairwise Pearson correlation between the row and the column variable.

Recent SNB Working Papers

2018-11 Maxime Phillot, Rina Rosenblatt-Wisch: Inflation Expectations: The Effect of Question Ordering on Forecast Inconsistencies

2018-10 Alex Cukierman, Thomas Lustenberger: International Evidence on Professional Interest Rate Forecasts: The Impact of Forecasting Ability

2018-9 Thomas Nitschka: Did China's anti-corruption campaign affect the risk premium on stocks of global luxury goods firms?

2018-8 Andreas M. Fischer, Lucca Zachmann: Do the rich pay their taxes early?

2018-7 Basil Guggenheim, Mario Meichle and Thomas Nellen: Confederation debt management since 1970

2018-6 Lucas Marc Fuhrer, Basil Guggenheim and Matthias Jüttner: What do Swiss franc Libor futures really tell us?

2018-5 Barthélémy Bonadio, Andreas M. Fischer and Philip Sauré: The speed of exchange rate pass-through

2018-4 Simone Auer, Christian Friedrich, Maja Ganarin, Teodora Paligorova, and Pascal Towbin: International Monetary Policy Transmission trough Banks in Small Open Economies

2018-3 Romain Baeriswyl, Camille Cornand and Bruno Ziliotto: Observing and shaping the market: the dilemma of central banks

2018-2 Adriel Jost: Cultural Differences in Monetary Policy Preferences

2018-1 Reto Foellmi, Adrian Jaeggi and Rina Rosenblatt-Wisch: Loss Aversion at the Aggregate Level Across Countries and its Relation to Economic Fundamentals

2017-18 Gregor Bäurle, Sarah M. Lein and Elizabeth Steiner: Employment Adjustment and Financial Constraints – Evidence from Firm-level Data

2017-17 Thomas Lustenberger and Enzo Rossi: The Social Value of Information: A Test of a Beauty and Non-Beauty Contest

2017-16 Aleksander Berentsen and Benjamin Müller: A Tale of Fire-Sales and Liquidity Hoarding

2017-15 Adriel Jost: Is Monetary Policy Too Complex for the Public? Evidence from the UK

2017-14 David R. Haab, Thomas Nitschka: Predicting returns on asset markets of a small, open economy and the influence of global risks

2017-13 Mathieu Grobéty: Government Debt and Growth: The Role of Liquidity

2017-12 Thomas Lustenberger and Enzo Rossi: Does Central Bank Transparency and Communication Affect Financial and Macroeconomic Forecasts?

2017-11 Andreas M. Fischer, Rafael Greminger and Christian Grisse: Portfolio rebalancing in times of stress

2017-10 Christian Grisse and Silvio Schumacher: The response of long-term yields to negative interest rates: evidence from Switzerland

SCHWEIZERISCHE NATIONALBANK
BANQUE NATIONALE SUISSE
BANCA NAZIONALE SVIZZERA
BANCA NAZIUNALA SVIZRA
SWISS NATIONAL BANK