Essays on Survey Expectations

Zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

von der KIT-Fakultät für
Wirtschaftswissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

M.Sc. Lora Pavlova

Tag der mündlichen Prüfung: 25.07.2023

Referent: Prof. Dr. Fabian Krüger
Korreferent: Prof. Dr. Christian Conrad

Karlsruhe, 2023

Acknowledgements

First and foremost I would like to express my deepest gratitude to Prof. Dr. Fabian Krüger, who has been my mentor for years, even prior to my Ph.D. Especially during my years as a PhD-student, I have received immense support and guidance from you and for that, I am very grateful. I would also like to thank Prof. Dr. Christian Conrad for inspiring me about research early on, in my years as a Master's student, and who has since continuously guided and supported me. I would like to extend my sincere thanks to Prof. Dr. Melanie Schienle, for always being there for me with helpful advice.

I would like to thank my co-authors and colleagues Mathias Hoffmann, Malte Knüppel, Emanuel Mönch, and Guido Schultefrankenfeld for our exciting and fruitful work. I would also like to thank all my colleagues at KIT for making my time there most enjoyable. Especially, I cannot thank Konstantin Görgen and Lotta Rütter enough, not only for making my transition from Heidelberg much easier but also for motivating me continuously throughout my research journey.

I am especially grateful to my family for supporting me to pursue a PhD-degree in the first place. Finally, I cannot express my gratitude to Jan Schmitt, who has believed in me and unconditionally supported me from the start and in the most difficult of times. This would not have been possible without you.

Contents

List of Figures

List of Tables

List of Abbreviations

AIT Average inflation targeting
AR Autoregressive model
ARMA Autoregressive Moving Average
ASA American Statistical Association
B Uncertainty index by Binder (2017)
BEA Bureau of Economic Analysis
BOP-HH Bundesbank Online Panel (Households)
BS Brier score
CBOE Chicago Board Options Exchange
CCPI Core Consumer Price Index
CDF Cumulative distribution function
CEPR Center for Economic and Policy Research
CES Consumer Expectations Survey
CPI Consumer Price Index
CRPS Continuous ranked probability score
CSCE Canadian Survey of Consumer Expectations
CY Current year
DIS Disagreement
DSGE Dynamic Stochastic General Equilibrium
EBS Expected Brier score
ECB European Central Bank
ECRPS Expected continuous ranked probability score
ELS Expected logarithmic score
EMW Engelberg et al. (2009) quantification method
EPU Economic Policy Uncertainty
ERPS Expected ranked probability score
FE Fixed event
Fed Federal Reserve Bank
FH Fixed horizon
FRBNY Federal Reserve Bank of New York
GDP Gross domestic product
HH Households
HICP Harmonized Index of Consumer Prices
IQR Interquartile range

IT Inflation Targeting
JLN Macro Uncertainty
LS Logarithmic score
MAM Mass-at-midpoint method
MSC Michigan Survey of Consumers
NBER National Bureau of Economic Research
NY Next year
OLS Ordinary Least Squares
PCE Personal Consumption Expenditures
PP Point forecast
RCT Randomized control trial
RGDP Real Gross Domestic Product
RMSE Root of the mean squared error
RPS Ranked probability score
RS Ex-post uncertainty
SCE Survey of Consumer Expectations
SD Standard deviation
SPF Survey of Professional Forecasters
US SPF US Survey of Professional Forecasters
VAR Vector autoregression
VIX Chicago Board Options Exchange's Volatility Index

1 Introduction

Expectations - the beliefs people hold about the course of future events - play a central role in economics. Especially when individuals are faced with choices under uncertainty, that is, for example, when the true probability of an occurrence is unknown, the subjective probabilities they hold about this event can be informative of their actions today. Recognizing their importance beyond pure behavioral data (Katona and Likert, 1946), researchers have started collecting subjective expectations data through surveys on a larger scale as early as the 1940s. While the use of the data remained in the background and was often disregarded for several decades afterwards, largely due to the rational expectations revolution (Muth, 1961), it resurfaced in the late 20th century.

The collection of expectations data has evolved greatly over time, with individuals first reporting simple qualitative forecasts to later answering more complex questions about numeric point predictions or even providing the whole subjective distribution for the variable of interest. Eliciting a person's probability distribution, or in other words her *subjective histogram*, is advantageous for several reasons: among others, it allows for computing a measure for the ex-ante uncertainty of the respondent. Charles Manski, one of the main advocates of the use of subjective expectations data in economics, and probabilistic ones in particular stresses that researchers 'should do more than take such measurements at face value' (Manksi, 2023, p. 809). He calls for a deeper understanding of the respondent's reporting behavior and potential changes in the latter due to different measurements or the availability of new information (Manksi, 2023). Even when using the simplest wording and question design, the elicitation of probabilistic expectations is not straightforward. Some characteristics of the survey responses, such as rounding and focal point responses (Hurd, 2009; Fischhoff and Bruine de Bruin, 1999; Manski and Molinari, 2010) pose challenges for the initial analysis as well as subsequent processing of the data. In the context of households, a wide range of topics is collected such as expectations relevant for savings and retirement, the labor and stock market, or expectations about house prices, growth, and inflation. As the management of the latter has become central to the conduct of monetary policy, multiple central banks have started collecting detailed data on the public's inflation expectations and are closely monitoring them, gauging risks for price stability. One possibility to steer inflation expectations of economic agents is to directly communicate with them. Expectation surveys offer policy-makers an efficient way to improve their strategic communication with the public by testing out how respondents' inflation expectations react to different types of central bank communication in experiments embedded in these surveys. More precisely, by providing survey participants with information e.g. about the central bank's target, instruments, or its strategy as a whole in a randomized control trial set-up, one can causally identify their effects on inflation expectations. See for example Coibion et al. (2022a, 2021, 2022b).

In this thesis, I investigate several of the topics mentioned above in greater detail. Conceptually, the thesis consists of three parts: (i) elicitation of consumer expectations in a probabilistic format, (ii) quantifying uncertainty at the individual level and at different horizons, (iii) application of consumer inflation expectations data collected through surveys to answer topical monetary policy questions. While Chapter 2 critically discusses different ways of eliciting probabilistic inflation expectations of consumers, Chapters 3 and 4 tackle the measurement of subjective uncertainty derived from such histograms. Chapter 5 illustrates how data on inflation expectations of consumers elicited in an information-experimental framework can be used to analyze how consumers would react to a hypothetical monetary policy strategy change in the euro area.

More precisely, the four chapters of this thesis contribute to the existing literature in the following ways. In Chapter 2 I briefly review the evolution of eliciting consumer inflation expectations via surveys. In particular, the focus is set on the role of question wording and format and how variations in those may result in variations in the data. Building upon insights gained from existing studies on survey design, I develop a simple experiment aimed at identifying the effects of changes in wording and format on consumers' mean expected inflation and the accompanying uncertainty. The experiment was pre-registered at the American Economic Association RCT Registry at `https://www.socialscienceregistry.org/trials/6482` and conducted as part of the Bundesbank Online Panel-Households (BOP-HH), an online consumer expectation panel lead by the Deutsche Bundesbank, in September 2020. I compare data on consumer inflation expectations from the default probabilistic format, widely utilized in multiple large-scale consumer surveys, to one from an alternative, simpler format, both collected via a randomized control trial (RCT). In a randomized fashion, participants receive variations in the question wording as well - they are asked either about the expected *inflation rate* or the expected change in *prices in general*. Additional to reporting the effects for mean expectations and uncertainty, I compare the different treatments across several dimensions such as anchoring of expectations, consistency, rounding, and non-response rates. The experimental results show evidence of rather strong framing effects on the elicited expectations data.

Building on the notion established in the first part of the thesis that probabilistic expectations can be informative of the underlying uncertainty, Chapter 3 proposes a new method for quantifying subjective uncertainty in expectations data. While there is a long history of asking professionals for their probabilistic forecast, in the context of households this is a relatively new phenomenon, implemented on a larger scale since the 2010s. To compute a person's subjective uncertainty, the standard methodology proposes fitting a flexible distribution to the discrete histograms of the respondents (Engelberg et al., 2009). Then the researcher can compute the desired measure of spread such as the standard deviation (SD) or the interquartile range (IQR) of the distribution which serves as a proxy for individual uncertainty. Since the standard methodology was developed for data collected from professionals, it does not take into account some characteristics typical for consumer data. This is an important aspect, as households often substantially differ from professionals in their reporting behavior, e.g. by reporting positive probability in one or two bins only, in outer bins, or in disjoint regions. Instead, the chapter proposes a measure of uncertainty based on the concept of entropy. More precisely, to use the entropy function associated with a scoring rule, the ranked probability score (Epstein, 1969), as a measure for uncertainty (ERPS). The ERPS is particularly attractive due to its simplicity, lack of restrictive assumptions on the underlying support of the histogram, and robustness. The latter is illustrated via several simulation studies.

Chapter 4 is also concerned with the measurement of uncertainty but in the context of professional forecasters. More precisely, it assesses the ex ante uncertainty for the US Survey of Professional Forecasters (US SPF), a quarterly panel survey of professionals widely utilized in the forecasting literature. In this survey, as well as in its European counterpart, the ECB SPF, participants are asked to assign probabilities to mutually exclusive intervals that cover different outcome ranges for growth, inflation, and unemployment variables, among others. In each quarter, respondents are asked to make a forecast for the current and next year (fixed-event forecasts). As one moves through the year the forecast horizon declines, causing a seasonal pattern in the dispersion of the histogram that can be problematic in subsequent analysis. The aim of the chapter is to derive a time series of fixed-horizon, e.g. 12-months ahead uncertainty, by combining the existing current- and next-year forecasts. During the process, several major issues are addressed, among others structural breaks in uncertainty which seem to stem from changes to the underlying response scale ('histogram bins').

Finally, motivated by the strategy review the European Central Bank (ECB) conducted prior to the strategic changes implemented as of July 2021[1], Chapter 5 tries to answer how consumers in Germany would react to a change in the ECB's mandate, by evaluating their reported probabilistic inflation expectations. The latter are elicited in an experimental framework from multiple waves of the BOP-HH in the autumn of 2020 and spring of 2021. In a randomized control trial (RCT), participants are shown information on the previous strategy of the ECB of aiming for an inflation rate 'close to but below 2% in the medium term', and on a hypothetical strategy similar to the one pursued by the Fed, namely flexible average inflation targeting (AIT) of maintaining inflation at 2% on average. In the experiment, the mechanism of the history-dependent regime is explained in layman's terms - if inflation were to fall below 2% for a particular amount of time, then subsequently the central bank would allow for it to exceed the target for some time, such that the target is met on average. Eliciting expectations prior to and after the information provision allows for identifying the causal effect of a change in the central banks' strategy on consumers' mean expected inflation over different horizons. The results suggest that households understand the stabilizing mechanism of the hypothetical strategy and adjust

[1] `https://www.ecb.europa.eu/home/search/review/html/index.en.html`

their expectations accordingly. That is, following a period of below-target inflation, consumers raise their inflation expectations under the alternative strategy.

While Chapter 2 is single-authored, Chapter 3 is joint work with Fabian Krüger and has been recently published in the *International Journal of Forecasting*, (2023)[2]. Chapter 4 is co-authored together with Malte Knüppel. Chapter 5 is joint work with Mathias Hoffmann, Emanuel Mönch, and Guido Schultefrankenfeld, and has been published at the *Journal of Monetary Economics*, Vol. 129 (2022), S52–S66[3]. The works have been presented at several conferences such as the Joint Deutsche Bundesbank-Banque de France Conference on Household Expectations, IWH-CIREQ-GW Macroeconometric Workshop: Micro Data and Macro Questions, Second Joint ECB-FRBNY Conference on Expectations Surveys, ESCB Research Cluster on Monetary Policy Workshop, and First Bergamo Workshop in Econometrics and Statistics.

[2] In press, corrected proof available online at `https://doi.org/10.1016/j.ijforecast.2023.06.001`.
[3] Original publication available at `https://doi.org/10.1016/j.jmoneco.2022.02.006`.

2 Framing Effects in Consumer Expectation Surveys

2.1 Introduction

In times of economic turmoil, it has become increasingly important that policymakers receive the most reliable and timely information on consumers' expectations about the future development of both macroeconomic and idiosyncratic outcomes. In the context of the broader public, one of the most effective ways of eliciting expectations in economics is directly surveying consumers. Such surveys are conducted by multiple central banks and research institutions, and this practice has become more widespread in recent years[1], with the focus shifting from point expectations to probabilistic ones. One particularly important example is inflation. Expectations about future price changes play a major role in forecasting inflation outcomes and economic activity, as well as in the transmission mechanism of monetary policy. Policymakers also carefully monitor inflation expectations for early signs of erosion of central bank credibility and de-anchoring of expectations that can lead to dangerous spirals (Nagel, 2022).

Manski highlights the 'importance of careful attention to question wording when eliciting expectations' (Manski, 2018, p. 440). For decades, it has been common practice to ask consumers to state a value for the expected 'change in prices in general'. A leading example is the Reuters/ University of Michigan Survey of Consumers (henceforth Michigan Survey), which elicits consumer expectations about price changes in the form of point forecasts since the late 1970s (Curtin, 1996). In a more recent survey, the Federal Reserve Bank of New York (FRBNY) adopted an alternative question wording that

[1] For a detailed list of consumer and firm surveys and the institutions conducting them see e.g. Weber et al. (2022), Coibion et al. (2020).

directly asks consumers about their expected inflation rate. This change aims for a more precise formulation in order to reduce ambiguity and variation in responses which stem from different interpretations rather than true differences in beliefs (Armantier et al., 2017). Furthermore, studies show that when confronted with the *inflation rate* formulation, consumers tend to think more often about the overall inflation rate rather than specific goods' prices (Bruine de Bruin et al., 2012). However, this comes at the cost of a potentially higher degree of complexity. Therefore, the effect of a change in wording might not be symmetric across demographic groups. While more educated and financially literate people may adjust their answers to better conform with economic terminology, for their counterparts this might lead to a higher item non-response rate (D'Acunto et al., 2020).

Together with a new wording, the FRBNY also introduced a new type of question to its survey, previously mostly utilized in surveys of professional forecasters such as the ECB and US SPF. While point predictions provide a view of the inflation rate expected 'on average', they contain no information about the individual uncertainty associated with the forecast. One possibility to capture this uncertainty is to elicit the whole subjective distribution by asking respondents to assign probabilities to a set of non-overlapping intervals representing possible outcome ranges of the forecasted variable (subjective histogram). While highly informative, this format also suffers several drawbacks. For once, it strongly relies on the respondent's capabilities of expressing herself via probabilities, which in turn requires a certain degree of numeracy and sophistication. For example, in the context of open-ended probability questions, studies document a high proportion of 50% answers as well as an over-proportional usage of 0 and 100%, also known as 'focal point responses' (Hurd, 2009; Dominitz and Manski, 1997). Stating 50% might indicate high epistemic uncertainty among respondents and signal that they struggle with numbers and probabilities, or to form a distribution about expected outcomes of an unknown concept (Fischhoff and Bruine de Bruin (1999), Bruine de Bruin et al. (2000)). Even if people have precise probabilities about future outcomes in mind, they may resort to rounding to facilitate communication (Manski and Molinari, 2010). The challenge for the researcher remains to distinguish between the two (for a more detailed discussion see Manski (2018) and references therein). A more recent study by Becker et al. (2023)

provides some experimental evidence of the large influence of the underlying response scale on the reported distribution of probabilities and subsequently derived inflation expectations and uncertainty measures.

This paper contributes to the strand of literature focusing on survey design. In the context of consumer (or firm) inflation expectations, several aspects have been addressed so far. Bruine de Bruin et al. (2017) discuss the effect of administration mode (e.g. face-to-face vs web-based surveys) and opportunities to revise answers similar to the practice in the Michigan Survey of Consumers (Curtin, 1996). Bruine de Bruin et al. (2012) and Bruine de Bruin et al. (2017) study, among others, the effect of wording on the central tendency and disagreement of inflation expectations and perceptions, and report differences between wording versions such as the 'inflation rate', 'prices in general' or 'prices you pay'. Dräger and Fritsche (2013) investigate similar aspects using representative data for the population of the city of Hamburg, Germany, and are, to a large extent, able to replicate previous findings for point expectations and perceptions. Coibion et al. (2020) do not find any systematic biases in the first and second moments based on variations in wording for firms in New Zealand. Phillot and Rosenblatt-Wisch (2018) examine the effect of question ordering on the forecast consistency of firms and report significant differences depending on whether the point or the density forecast is asked first. Becker et al. (2023) focus exclusively on how the response scale shapes expectations derived from subjective histograms of households and provide some evidence for sensitivity to the underlying bin definitions. Another recent paper closely related to the current study is the one by Hayo and Méon (2022), in which the authors assess the effect of guided vs. non-guided questions on the reported inflation expectations and non-response rates.

Using a nationally representative consumer expectations survey in Germany, this paper estimates the causal effects of question wording and format on the reported inflation expectations, both jointly and separately. To the best of the author's knowledge, this is the first study to examine these aspects in the context of consumer probabilistic expectations for Germany. I find that presenting respondents with an alternative format asking for the maximum, minimum, and mode of their distribution, leads to a slight increase in non-response rates. In addition, mean expected inflation derived from the probabilistic forecast increases by more than 1 pp. On the contrary, the individual uncertainty of the respondent measured by the standard deviation of her distribution, is estimated to be 0.6 to 1 pp smaller, suggesting that standard assumptions on the endpoints of the outer intervals might be less precise. As to the effect of wording on reported short-run inflation expectations, I can largely confirm findings from previous literature that the expectations of those provided with the *inflation rate* wording seem less upward-biased and more concentrated around 2%. However, a substantial share of respondents states they think most about specific goods' prices such as those of food and gas, independent of wording choice. Even when asked directly about expected inflation, only about one-quarter of consumers report actually thinking about it when producing their forecast.

The paper is structured as follows. Section 2.2 discusses the role of wording and format in the elicitation of consumer inflation expectations in greater detail. Section 2.3 presents the experimental framework and Section 2.4 gives an overview of the data set. Section 2.5 provides descriptive statistics about the collected expectations data, while Section 2.6 summarizes the estimated treatment effects of wording and format. Section 2.7 concludes.

2.2 The Role of Wording and Format in Eliciting Households' Inflation Expectations

Ultimately, survey questions should be interpreted in the same way by both researchers and respondents (Bruine de Bruin et al., 2011). To avoid miscommunication and diffuse interpretations, question wording should be as precise as possible. However, economic terms such as *'rate of inflation'* or *'Consumer Price Index'* might not be understandable for all respondents and their inclusion in the question might cause a higher rate of non-response, especially among groups with lower education and financial literacy, or limited cognitive abilities (D'Acunto et al., 2020).

For decades people surveyed in the Michigan Survey were asked for their expectations about the 'change in prices in general'. While simpler in design, the answers to this question have displayed large disagreement and frequently exhibited extreme values (Candia et al., 2020), overstating realized inflation despite follow-up probing and truncation of outliers (D'Acunto et al., 2023). Variation in inflation expectations can be, at least partially, attributed to factors related to personal experiences: differences in consumption baskets and thus exposure to different prices (D'Acunto et al., 2020), cohorts living through various inflation regimes (Malmendier and Nagel (2016), Goldfayn-Frank and Wohlfart (2020)), diverse financial planning horizons (Bruine de Bruin et al., 2010). Apart from varying experiences, people might differ in the process of forming expectations (Bruine de Bruin et al., 2010). That is, whether they think about person-specific financial experiences or price changes on a broader scale such as those reflected in the CPI index. Bruine de Bruin et al. (2010) report that people tend to interpret the *'prices in general'* wording as asking most for the former. Overall, the authors find that thinking about personal experiences is associated with providing more extreme values for expected inflation, which might explain the observed upward bias.

Indeed, previous experimental studies have shown that question wording significantly impacts the resulting point forecasts. Using a representative US sample, Bruine de Bruin et al. (2012) document that both means and medians of the aggregate point predictions distribution from a question using a 'prices in general' or 'prices you pay'-wording exceed those produced by an 'inflation rate'-wording. This also holds true for respondents' perceptions of inflation. While the authors find almost equal non-response rates across wordings, they do report participants having more difficulty understanding terms such as the 'inflation rate'. For Europe, Bruine de Bruin et al. (2017) document similar findings using a nationally representative survey of Dutch consumers, albeit the differences in levels between various wordings appear somewhat less pronounced[2].

While the Michigan Survey has asked respondents about the expected rate of price changes, or in other words a point prediction, another important component of expectations is their uncertainty. For example, one can compute the *disagreement* across forecasters from the standard deviation of their point predictions, which provides some notion of aggregate uncertainty (Glas, 2020). However, if interested in individual uncertainty, one could elicit the participant's whole subjective distribution of possible inflation

[2] Table 4 of Bruine de Bruin et al. (2017) shows that 'prices in general' wording does not produce significantly higher inflation expectations for initial responses in a web mode. However, one should also note that the data from the US experiment of Bruine de Bruin et al. (2012) was elicited during a period with high observed inflation, which can additionally exacerbate differences across wordings.

outcomes over the horizon of interest. One of the pioneers in this field is the Federal Reserve Bank of New York (FRBNY). In 2013, the FRBNY launched a large-scale consumer survey on expectations - 'The Survey of Consumer Expectations' (henceforth SCE) - and introduced a probabilistic question format for the rate of inflation, average home price, and personal wage over different horizons (Armantier et al., 2017). For example, the question about short-term inflation expectations is as follows[3]:

Q9 Now we would like you to think about the different things that may happen to inflation over the next 12 months. [...] In your view, what would you say is the percent chance that over the next 12 months...

the rate of inflation will be 12% or higher

the rate of inflation will be between 8% and 12%

the rate of inflation will be between 4% and 8%

the rate of inflation will be between 2% and 4%

the rate of inflation will be between 0% and 2%

the rate of deflation (opposite of inflation) will be between 0% and 2%

the rate of deflation (opposite of inflation) will be between 2% and 4%

the rate of deflation (opposite of inflation) will be between 4% and 8%

the rate of deflation (opposite of inflation) will be between 8% and 12%

the rate of deflation (opposite of inflation) will be 12% or higher

In essence, based on their personal assessment respondents are asked to fill out probabilities to different ranges ('bins') of inflation outcomes such that they sum up to 100%. Such a type of question, previously mostly used in surveys of professional forecasters, allows for computing a measure of individual uncertainty such as the variance or IQR or some other measure of the spread after fitting a continuous distribution to

[3] All questions from the SCE questionnaire are available online at `https://www.newyorkfed.org/microeconomics/databank.html`.

the discrete histogram (Engelberg et al., 2009)[4]. While highly informative, the standard method of quantification needed to compute an uncertainty measure often imposes rather restrictive assumptions on the underlying subjective distribution and is sensitive to small changes in the reported probabilities (Krüger and Pavlova, 2021).

Furthermore, another important dimension is that the resulting data highly depends on both people's preference to think and their ability to express their expectations and beliefs using numerical probabilities (Manski, 2018). While an overall willingness to convey expectations in a probabilistic manner has been exhibited by respondents (Armantier et al., 2013), several artifacts have been observed in the data so far. For instance, Bruine de Bruin et al. (2000) find that when asked to estimate the risk of an event happening and report the corresponding probability, uncertain respondents often resort to using the number 50. The answer can be a sign that the participant struggles to express her feeling as a number or experiences high epistemic uncertainty, rather than an intended use of 50% (Fischhoff and Bruine de Bruin, 1999). Moreover, this pattern is heterogeneous with respect to age, education, and numeracy (Bruine de Bruin et al., 2000). More generally, the usage of round numbers when reporting inflation expectations is found to be a proxy for subjective uncertainty (Binder, 2017), but can also be used by consumers to facilitate communication (Manski, 2018). Based on the raw data alone, it is unclear which one applies. Manski and Molinari (2010) suggest examining the set of responses provided by the survey participant in order to draw on her rounding pattern, i.e. exact or rounded reporting. However, people might adopt different patterns depending on whether the elicited outcome is personal or macroeconomic, and if the latter - how familiar they are with the concept, and so on.

In a more recent work, Becker et al. (2023) investigate the effects of changes in the response scale on density forecasts. The authors show that shifting or compressing the scale affects forecasts accordingly and thus question the suitability of subjective histograms to evaluate whether inflation expectations of the broader public are anchored or not. Moreover, even if policymakers are more interested in expectations volatility rather than their level since consumers are well-known to strongly overestimate inflation

[4] An alternative approach, also known as 'mass-at-midpoint' assumes that the probability mass is placed on the midpoint of each interval and also allows for computing the measures mentioned above (Glas and Hartmann, 2022).

even in times when it is low and stable, uncertainty derived from density forecasts also seems very prone to changes as the underlying intervals change (Becker et al., 2023). Even though there are clear benefits from using a probabilistic question and it has since been adopted by multiple central banks and research institutions in their consumer surveys, these aspects call for examining a viable alternative to this format.

2.3 A Randomized Experiment among Consumers in Germany

The current study aims to test the effect of a change in wording and format on the responses about short-term inflation expectations of a representative sample of consumers in Germany[5]. To this end, the survey participants are randomly split into four treatment arms and presented with different versions of a question asking for their inflation expectations. The treatments are as follows:

- Group A1: probabilistic question about the inflation rate (henceforth default)
- Group B1: min-max question about the inflation rate
- Group C1: probabilistic question about changes in prices in general
- Group D1: min-max question about changes in prices in general

To test the effects of wording, I replicate the experimental setup from Bruine de Bruin et al. (2010) and Bruine de Bruin et al. (2012) and extend it by assessing the effects in the context of probabilistic expectations. Each two sub-samples receive the *inflation rate* (A1, B1) or *change in prices in general* (C1, D1) wording. Earlier experiments such as Bruine de Bruin et al. (2012) consider a third alternative, namely 'prices you pay'. For this, however, the same arguments as for the *prices in general* apply, and in most cases, it does not produce significantly different expectations, which is why I do not include it in the experiment. Based on insights gained by existing research, I hypothesize that a more straightforward wording such as *prices in general* generates higher expected

[5]The trial was pre-registered at the American Economic Association RCT Registry at `https://www.socialscienceregistry.org/trials/6482`

inflation and larger variation in individual responses. On the other hand, the *inflation rate* option, while prompting people to think more about price changes on a broader scale, thus reducing the upward bias in expectations, is more hard-to-read and leads to a higher non-response rate. So moving from an *inflation rate* formulation to one using *prices in general* instead, one should observe an upward shift in expectations coupled with higher uncertainty, but a lower non-response rate. One can estimate the treatment effects of wording in a simple linear regression framework as described in Equation (1):

$$y_i = \alpha + \beta \text{prices}_i + \gamma \text{min-max}_i + \delta \text{joint}_i + \varepsilon_i, \tag{1}$$

where y_i is the inflation expectations measure of interest of person i. In the study, I will focus mostly on measures of central tendency and uncertainty such as $\mathrm{E}[\pi_{\mathrm{t+12}}]$ and $\sigma^{\pi_{t+12}}$ derived from the reported subjective histograms. α represents the respective measure for group A1 which receives the default version of the question as described in Section 2.2. The remaining terms on the right-hand side capture expected inflation (or uncertainty) for the other three treatment groups relative to the default group. That is, the variables *prices*$_i$, *min-max*$_i$ and *joint*$_i$ are dummies taking unit value when respondent i is assigned to group B1, C1 or D1, respectively. As each survey participant is randomly sub-sampled into only one treatment group, the treatment dummies cannot have a value of one simultaneously. Conditional on the probabilistic format, the hypothesis discussed above implies a positive sign on β in Equation (1), that is the simpler *prices in general* wording would increase mean expected inflation as well as the respondent's uncertainty.

Secondly, I investigate the differences between two different approaches to eliciting people's subjective distribution about the inflation rate over the next 12 months on the resulting data. While groups A1 and C1 receive the standard probabilistic question, the remaining two groups are given a formulation that is free of probabilities. The alternative question implemented in the experiment, which I will henceforth refer to as 'min-max', is the following:

Question 1: What do you think the rate of inflation (or rate of deflation) is most likely to be over the next twelve months? What will the rate of inflation be as a maximum or minimum value?

For group C1 the wording is adjusted accordingly. The idea behind this treatment is to compare the subjective distributions which result from the two question formats and also how they interact with variations in the wording. While the question above is more limited regarding the functional form of the distribution, it has the advantage that respondents can explicitly define the ends of the support of their histogram making additional assumptions unnecessary. The fact that they are not presented with any numbers or ranges beforehand reduces the possibility of respondents altering their answers to better fit the current setting, e.g. by placing probability symmetrically around zero; placing more probability on lower numerical values if their point prediction lies in an outer bin. For instance, Becker et al. (2023) show that participants tend to assign higher probability mass to a given numeric range, as the number of bins representing this range, increases. This phenomenon observed in an experimental setup has the potential to overstate the degree of anchoring to the inflation target or any other value for that matter, depending on how the bin definitions are specified. By design, Question 1 also reduces the share of 'problematic cases' in the data such as responses that include one or more disjoint regions with positive probability mass or exhibit bi-modality. While such cases are not dominant in household expectations data, they do represent a non-trivial share[6] and are more difficult to handle by standard methodology. Overall, the degree of complexity of the question is reduced and the information respondents receive beforehand is minimized.

Questions similar in design have been implemented in the context of households (Coibion et al., 2022a; Christelis et al., 2020) and firms (Altig et al., 2022), however, in either case, the respondents are asked to additionally assign a probability to the corresponding scenario, e.g to the 'lowest' or 'highest' outcome, or the average of the two. In contrast, the setup proposed above, which is closely related to the one used for eliciting a firm's expected sales growth in the ifo Business Survey (Bachmann et al., 2021), is completely free from numerical probabilities. Huisman et al. (2021) also use a similar framework, asking for a point forecast, a minimum, and a maximum level of the AEA index to collect stock market expectations of Dutch investors.

Ideally, in a scenario where wording and format do not influence response behavior, the resulting distributions of the four question versions should be identical. However, I

[6] For the SCE between June 2013 and November 2020, Zhao (2022) documents less than 5% of histograms with disjoint regions and about 14% containing multiple modes.

predict that while more precise in design, a pure probabilistic question also: (i) generates a higher non-response rate and (ii) causes respondents to place more probability mass in the middle intervals. I conjecture that moving from a probabilistic formulation with predefined intervals to a less restrictive one, one should observe an upward shift in expectations and a larger variation in the individual responses, or put differently a positive γ in Equation (1).

Finally, the coefficient δ on the fourth term in Equation (1) captures the joint influence of both simpler wording and format combined in the final treatment. Given the separate effects described above, δ is expected to have a positive sign. If δ significantly exceeds the sum of the other two coefficients β and γ, this will provide evidence for an additional interaction effect between wording and format beyond the individual effects identified before. To summarize, for both measures of central tendency such as means or medians as well as for uncertainty measured by the individual standard deviation, I hypothesize the following:

- wording increases both the level and spread in inflation expectations or $\beta > 0$
- format also increases both, $\gamma > 0$
- consequently, $\delta > 0$.

In addition to the statements above, I examine whether the effect of different question wording and format is symmetric across demographic groups. Given findings in previous studies for point forecasts (Dräger and Fritsche, 2013), a certain degree of asymmetry is to be expected. While under the *inflation rate* wording the expectations of better-educated respondents with higher financial literacy would be close to realized rates, those of their counterparts might remain biased as they experience more difficulty interpreting the question or are not aware of the concept. In the RCT setup, this would imply that β would be larger for the latter groups. Intuitively, a change in wording would be effective in reducing the bias in expectations for those who understand the economic terminology behind it. For the effect of format, I hypothesize the opposite pattern. Conditional on higher financial literacy, the anchoring effect of providing respondents with predefined intervals should be smaller. That is, γ should be smaller for these groups than the estimate for less financially literate populations.

Furthermore, to shed more light on whether wording affects the variability of responses through the 'question-interpretation' channel, respondents are asked what they thought most about when producing their forecast. The question is adapted from Bruine de Bruin et al. (2010), where it was successfully implemented as part of a web-based survey among members of the RAND's American Life Panel. Respondents are presented with five options they can choose from, whereby one option is open-ended. The options selected for the current experiment were the top-rated in the original one (see Bruine de Bruin et al. (2010), Table 2). The question is as follows:

Question 2: What did you think about most when answering the question about your inflation rate expectations before?[7]

- Prices you pay in your everyday life such as food and gasoline
- Prices Germans pay
- Germany's inflation rate
- Changes in the cost of living
- Other specific prices (*please name*)

One can expect that people who were randomly assigned to the *prices in general* treatment arm, should be more likely to select 'everyday prices'. Accordingly, if interpreted correctly, respondents from the *inflation rate* treatment group, should most often select the corresponding option - 'Germany's inflation rate'. Furthermore, it would be interesting to determine whether there is any systematic heterogeneity by demographic characteristics, e.g. whether populations with lower financial literacy think more often of personal experiences, while simultaneously controlling for a wording effect. In the remainder of the paper, I try to answer these questions.

[7] Again, the wording is adjusted accordingly for groups C1 and D1. For the exact question wording see Appendix 2.8.3.

2.4 Data Set

The experiment was conducted as part of the 9th wave of the Bundesbank Online Panel - Households (henceforth BOP-HH) in September 2020. BOP-HH is an online survey currently conducted at a monthly frequency, which involves a variety of topics on both personal and economic outcomes. Respondents are selected on a random basis from an offline recruited panel. The sample size varies from 2,000 to 5,000 respondents and contains a panel component. The survey is representative of the German online population of age 16 and above. In addition to the core and project-specific answers, information on the respondent's age, education, employment, income, household size and the number of children, and others, is also provided. For further information on the survey see Beckmann and Schmidt (2020).

In the several months preceding the September wave, Germany experienced low inflation rates[8], partly influenced by the temporary value-added tax (VAT) cut, which was in effect in the second half of 2020. This was reflected in the short-run inflation expectations of BOP-HH respondents, which after an upward shift at the beginning of the pandemic, gradually subsided to pre-pandemic levels in the autumn of 2020. While expectations remained relatively stable, uncertainty over the first year of the pandemic rose significantly[9]. Such a phenomenon was also observed in other consumer expectation surveys such as the SCE (Armantier et al., 2021).

The September 2020 wave of BOP-HH, in which the experiment was conducted, has a larger sample size of approximately 4,000 respondents. Thus, in each of the four treatment arms, there are about 1,000 participants. For the treatment groups who receive the *prices in general* option, the wording across all questions is adjusted accordingly. Respondents who receive the *inflation rate* versions are subject to an information treatment, established in previous waves of the survey. More precisely, they are shown a short definition of what inflation is: *'Inflation is the percentage increase in the general price level. It is mostly measured using the consumer price index. A decrease in the price level is generally*

[8]The monthly, year-on-year inflation rate for June, July, and August 2020 was at 0.8, 0.0, and −0.1. Source: Statistisches Bundesamt (Destatis).

[9]Summary statistics on inflation expectations and uncertainty are published monthly at `https://www.bundesbank.de/en/bundesbank/research/survey-on-consumer-expectations/inflation-expectations-848334`.

described as deflation.' As part of the core module of the survey, participants receive a number of questions specifically targeted at their inflation expectations[10] First, they are asked about their perception of the inflation rate development over the last 12 months. Then, regarding the upcoming 12 months in the following order respondents report whether they expect inflation or deflation, and if so, at how much percent (point prediction). Finally, depending on the treatment group, they are presented with either the probabilistic or the question asking for the mode, minimum and maximum. In the probabilistic question treatment, survey respondents have to assign probabilities to bins that the inflation rate (or change in prices in general) will fall into the range of the respective bin, analogously to the procedure described in Section 2.2. The bins are of different lengths, narrower and symmetric around zero, and the outer bins are open-ended. Bin definitions follow those from the SCE, albeit, neither the design of BOP-HH nor the SCE specifies which intervals contain the endpoints. To enable further processing of the data, I assume all intervals are right-closed, except the last one.

$$(-\infty, -12], (-12, -8], (-8, 4], (-4, -2], (-2, 0], (0, 2], (2, 4], (4, 8], (8, 12], (12, \infty)$$

Additionally, participants are instructed that the sum across probabilities should be zero. As they insert their answers, the current sum of the probabilities is displayed. If one attempts to skip the question, two options of non-response are shown: 'Don't know' and 'No answer'.

In the experiment, the other half of the participants is presented with the min-max question. In essence, by answering Question 1 the respondent reports both ends of the support of her subjective distribution, a and b, as well as the mode of the distribution, c. Using this information one can fit a simple triangular distribution to the subjective probabilities, which is a common practice in empirical literature[11]. Loosening the symmetry assumption on the form of the distribution allows for more flexibility compared to the standard practice. It follows that the mean and the variance of the distribution

[10] A detailed overview of the survey questions is documented in Appendix 2.8.3. Additionally, Deutsche Bundesbank makes all past questionnaires, including the one used in the September 2020 wave, available online at `https://www.bundesbank.de/en/bundesbank/research/survey-on-consumer-expectations/questionnaires-850746`.

[11] It is common to fit an isosceles triangle distribution, which is a special case of the triangular distribution (Engelberg et al., 2009)

are:

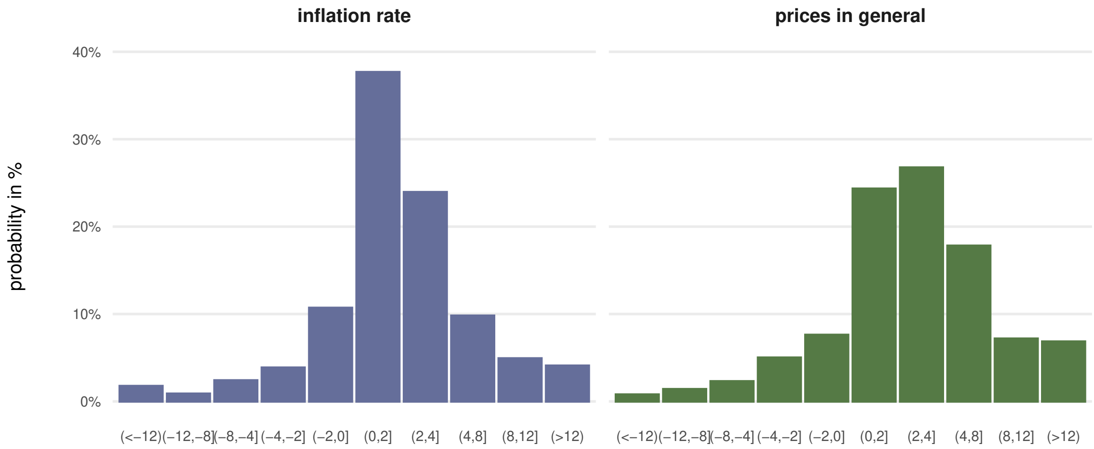
$$\begin{aligned} \mathrm{E}[\pi_{t+12}] &= \tfrac{a+b+c}{3} \\ \sigma^2_{\pi_{t+12}} &= \tfrac{a^2+b^2+c^2-ab-bc-ac}{18} \end{aligned}$$

For fitting a distribution to the subjective histograms produced by the probabilistic question I broadly follow Engelberg et al. (2009) and fit a symmetric triangular distribution to histograms with at most two bins with positive probability and a generalized Beta distribution else. Some necessary adjustments are made, given the fact that the method was initially designed for bins of equal width, which is not the case in the current setup[12]. Following Armantier et al. (2017) I impose an upper bound of the open intervals at ± 38. Excluding respondents who did not provide an answer leaves us with 1,892 responses for the probabilistic question and 1,794 for the min-max question. Generally, one would like to further leave out observations that either (i) do not sum up to 100% for the probabilistic question or (ii) have a reported mode outside of bounds in the min-max setup. Such cases of inconsistency do not occur in the data. In the following, I provide some descriptions and summary statistics on the elicited expectations data.

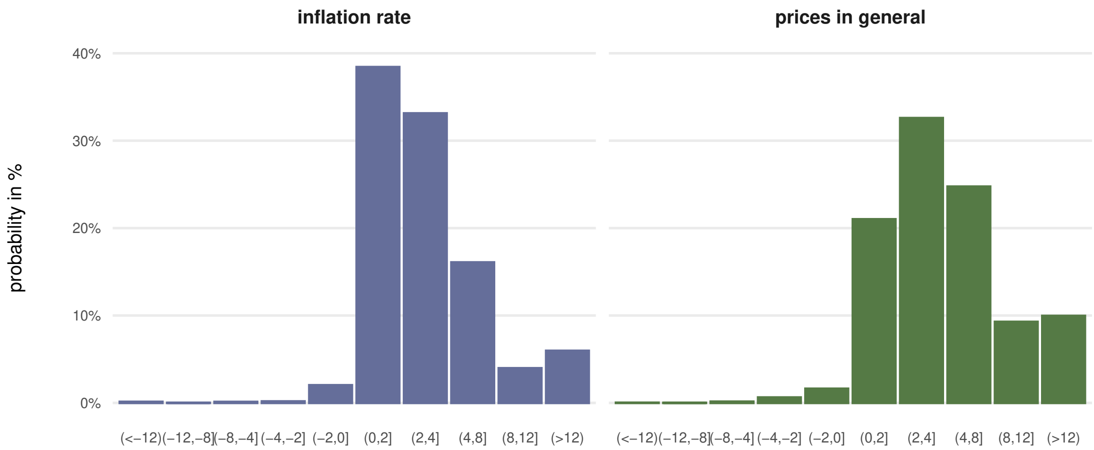
2.5 Descriptive Analysis

The upper panel of Figure 2.1 depicts the distributions of the respondents' probabilistic expectations. The left panel shows the subjective histograms for participants sub-sampled into the *inflation rate* treatment group, the right one for the *prices in general*. The distributions are somewhat similar in shape, with probability assigned predominantly to positive outcomes. The interval $(0, 4]$, which contains the official inflation target as the midpoint, attracts more probability in the *inflation rate* treatment than in the *prices in general* treatment - 62% vs 51%. With about one-third probability mass, respondents assigned to the latter group deem higher inflation outcomes ($> 4\%$) a lot more probable. In both cases, respondents consider deflationary trends quite likely - in the *inflation rate* slightly more than in the *prices in general*, with the difference being statistically

[12]See Krüger and Pavlova (2021) and the associated software available at `https://github.com/FK83/forecasthistogram`.



(a) probabilistic format (default)



(b) min-max format

Figure 2.1: Average subjective probabilities for different ranges for inflation from the probabilistic and min-max format. Left panels display probabilities assigned by those who received the *inflation rate* wording, right ones for those with the *prices in general* wording.

significant at the 1%-level. While for the *prices* wording, the probability of deflation gradually declines as the negative values become larger, for the counterpart, one can observe a slight peak in probability mass in the very first interval. This might indicate that some respondents simply do not understand the question and assign 100% to this bin in order to skip the question.

For comparability, the lower panel of Figure 2.1 reports the difference $F(u_k)_i - F(u_{k-1})_i$, where u is the upper bound of the interval k with $u \in \{-12, -8, -4, -2, 0, 2, 4, 8, 12, \infty\}$ and $F(x)_i$ is the CDF of the triangular distribution of respondent i, based on the reported parameters, a, b, and c from the min-max question. In a sense, the reported distributions are 'discretized' to match the intervals of the probabilistic format and then averaged over all respondents. The most striking feature of the data is what appears to be a zero lower bound on expectations. Similar observations have been documented in earlier studies e.g. by Blanchflower and MacCoille (2009) for the UK. Expectations that are predominantly positive are consistent with previous findings that positive price changes attract more of the consumers' attention and influence expectations more than negative ones (D'Acunto et al. (2021a), Cavallo et al. (2017)). The average probability assigned to negative inflation rates plummets from 19.6% to 2.4%, and 17.1% to 2.4% in each wording variation. This is in line with the findings of Becker et al. (2023) that providing respondents with more intervals representing negative values of inflation increases the probability placed in those. In the more extreme case of the min-max question where respondents are not confronted with any intervals, the probability of deflation dramatically declines. Apart from that, the distributions exhibit similar features with more probability mass concentrated in the right tail, especially for the *prices in general* wording. In the following, I discuss them in greater detail.

2.5.1 Item Non-response

Contrary to what was expected, the probabilistic question, despite its greater complexity, yields lower non-response rates as indicated in Table 2.1. With almost 7% of all answers, the item non-response is low and there are no differences across wordings. In the min-max setup, non-response rates are significantly higher for each wording. Even if we account for respondents who put 100% probability in the very first bin (0.6%), this would not

Table 2.1: Response patterns in the probabilistic and min-max questions

	probabilistic question		min-max question	
	inflation rate	*prices*	*inflation rate*	*prices*
Non-response				
item non-response	6.7	6.6	13.4^{m1}	$9.6^{p1,m5}$
Anchoring				
$P(0 < \pi^e \leq 2)$	37.7	24.3^{p1}	38.4	$21.0^{p1,m1}$
$P(\pi^e < 0)$	19.6	17.1^{p10}	2.5^{m1}	2.4^{m1}
Uncertainty and rouding				
sparse histogram	52.3	42.2^{p1}		
using 50-50% responses	7.7	7.1		
at least one outer bin	19.4	26.8^{p1}		
mean number of bins	3.2	3.5^{p1}		
Mode is multiple of 5			19.6	36.1^{p1}
Min is multiple of 5			16.3	24.9^{p1}
Max is multiple of 5			24.3	43.5^{p1}
Consistency				
PP not in support	11.0	15.1^{p5}	14.2^{m1}	$8.2^{p1,m1}$
$q_5 <$ PP $> q_{95}$	24.8	27.0	37.3^{m1}	$31.5^{p5,m5}$
$P((X_{pp}, Y_{pp}]) = 0$	16.3	15.6		
contain disjoint regions	3.3	2.4		
Observations	944	948	879	915

Note:^{p1}, ^{p5}, ^{p10} indicate that the corresponding measure is significantly different in the *prices in general* from the *inflation rate* wording at the 1, 5, and 10%-level.^{m1}, ^{m5}, ^{m10} indicate that the corresponding measure is significantly different in the *min-max* from the *probabilistic format* at the 1, 5, and 10%-level. Shares and probabilities are reported in percentage points. The reported differences are based on χ-squared, Wilcoxon-Mann-Whitney, or Kolmogorov-Smirnov tests, depending on the nature of the underlying variable. X_{pp} and Y_{pp} are the endpoints of the interval that contains the point prediction.

make up for the difference between formats. A recent study by Hayo and Méon (2022) provides a possible explanation for this result. In an experimental setup, the authors test the effects of guided vs. non-guided questions about inflation on non-response rates. In the non-guided version, respondents receive an open-ended question about a point forecast, whereas in the guided version they can choose one out of a number of pre-defined intervals. Hayo and Méon (2022) document significantly higher shares of non-response for the open-ended question and justify this result with the concept of 'social desirability bias'. That is, respondents evade answers they consider undesirable to the interviewer, researcher, or survey designer. While in an open-ended framework, this will result in an item non-response, in one with pre-defined answer options, respondents are 'guided' to an answer, they would otherwise have not provided (Hayo and Méon, 2022). The fact that significant differences between wordings do arise in the min-max setup, with *inflation rate* producing a higher non-response share than the *prices in general* by about 4 pp, additionally supports this argument. While participants experience more difficulty answering the *inflation rate* question, providing them with a guided question offsets the difference between wordings. Generally higher response rates are desirable in surveys, but as Hayo and Méon (2022) point out, such answers might simply increase the noise in the data and are less or not at all informative of the true beliefs of the respondent. This will especially be the case if the non-response rates are driven mostly by less educated, lower-income respondents in the min-max setup but not in the probabilistic one. To this end, in Table 2.5 in the Appendix, I report the average marginal effects from a probit regression of the dummy variable $d_{non-response}$, which takes unit value if there is a missing value, on socio-demographic characteristics. Columns (1) to (4) report these effects for the four treatment arms separately. The evidence appears mixed at best. One observes that older respondents are more reluctant to provide an answer as well as females. Also, lower-income respondents tend not to give a forecast more often, but the effects are roughly symmetric across the treatment arms except in Column (4), where instead we see some significant effects of household size and whether or not the respondent was living in East Germany pre-1989.

Additionally, an interesting phenomenon emerges in the data: in 46 occasions respondents provided the same value for the minimum, maximum, and mode. On the one hand, this example can be interpreted as the participant being absolutely certain of the inflation outcome. Alternatively, as the response is not informative of uncertainty, and if it does not coincide with the preceding point prediction, the data could be treated as missing. In any case, there should be a clearer approach on how to handle such problematic cases. Additional patterns with respect to uncertainty, consistency, and anchoring are documented in Table 2.1 and I discuss them in more detail below.

2.5.2 Anchoring of Expectations

Policy-makers often keep a close eye on developments in consumer expectations, in particular about inflation, looking for early signs of de-anchoring and gauging risks for price stability (Nagel, 2022). We can consider the probability assigned to the interval $(0\%, 2\%]$ as a crude measure for expectations being anchored[13]. Previous studies raise concerns about the default probabilistic format overstating the degree of anchoring of expectations, as it contains a larger number of intervals around zero (Becker et al., 2023). The collected data indeed provides some evidence supporting this notion: for the *prices in general* wording, the probabilistic question produces significantly more 'anchored' inflation expectations than the min-max setup, however, the difference of 3 pp is rather small. For the *inflation rate* wording, the question format does not appear to affect the probability assigned to inflation being close to the target. Overall, it seems that wording has a much larger anchoring effect, as within each format, in the *inflation rate* treatment $P(0 < \pi^e \leq 2)$ amounts to roughly 38%, or almost double to the one assigned in the counterpart treatments.

2.5.3 Rounding

It has long been acknowledged that rounding can convey some idea for the respondents' uncertainty. In the context of consumer inflation expectations, Binder (2017) first introduces a rounding-based uncertainty measure utilizing point forecasts from the Michigan Survey. A number of earlier studies document that survey participants often report 0, 50, or 100% when they are uncertain of a risk assessment (Fischhoff and Bruine de Bruin, 1999). Additionally, some respondents may resort to rounding so as to simplify communication (Manski, 2018). Following the latter intuition, rounding might be viewed as an undesirable feature of the data, as it makes interpretation more difficult.

[13] At the time the survey was conducted in September 2020, the ECB's inflation target was still defined as 'close to, but below 2%' see e.g. `https://www.ecb.europa.eu/home/search/review/html/price-stability-objective.en.html`

In a similar fashion to Binder (2017), in Table 2.1, I report the share of minimum, maximum, and mode that are multiples of five in the min-max question. Overall, a considerable share of respondents provides rounded values with substantial differences between wording choices: the *prices in general* wording produces much more uncertain responses, which is somewhat intuitive given the higher volatility of food and energy prices. Notably, it is the maximum value for expected inflation that is most likely a multiple of five, followed by the mode and minimum. Since, to the best of the author's knowledge, the min-max question has not yet been implemented in other large-scale consumer surveys that can be used for comparison, it is difficult to judge whether these shares are unusually high or not. The highest value observed in the data (43.5% for the maximum in group D1) is close to the one observed for point predictions in the Michigan Survey - 41.4% (see Table A.1 of Binder (2017)). I further compute a variable *round* $\in \{0, 1, 2, 3\}$ indicating if none, one, two, or all of the reported values in the min-max question are multiples of five. The estimates of a linear regression of the IQR as a proxy for individual uncertainty on *round*, each respondent's point forecast, and $m5$, a dummy indicating whether this forecast is also a multiple of 5 are displayed in Table 2.6. They suggest that respondents who report round values in the min-max question are significantly more uncertain than non-rounders. Notably, the variable *round* as well as the level of the point forecast appear to be the most robust predictors of uncertainty.

2.5.4 Bin Usage

More than half of the respondents sub-sampled into the *inflation rate* wording group, report histograms with positive probability in at most two bins (sparse histograms). The corresponding proportion for the *prices in general* treatment is lower at 42%. This is comparable with other consumer surveys using a probabilistic format such as the SCE (Krüger and Pavlova, 2021). Generally, the *prices in general* wording seems to prompt respondents to use a higher number of bins as depicted in Figure 2.2. While the number of used bins gradually declines in the left panel of Figure 2.2, with a final peak at 10, for the *prices in general* wording, three peaks are visible: at 1, 5, and 10. To quantify better the effect of wording on the number of bins with positive probability in Table 2.7 in the Appendix I report the estimates of a Poisson and quasi-Poisson regression of the number

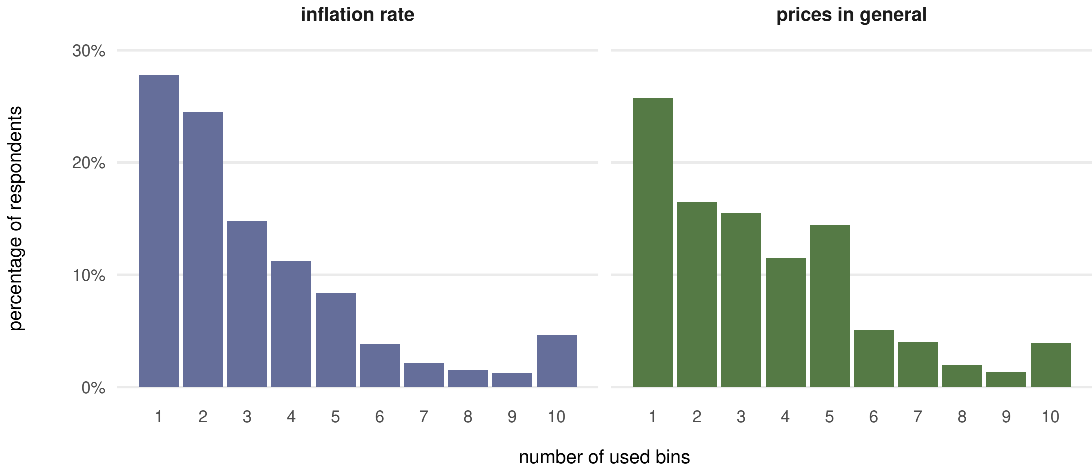


Figure 2.2: Number of bins used in the probabilistic question by wording

of used bins on the wording and a set of demographic controls. On average, receiving the *prices in general* formulation increases the number of bins with positive probability by 0.12 (0.13 respectively).

About one-fifth of the people who received the *inflation rate* treatment report positive probability in at least one outer bin, for the competing group this amounts to 26.8%. This result can be interpreted as being in line with previous literature which has found that when confronted with the *prices in general* wording, people tend to extrapolate based on their personal shopping experiences (Bruine de Bruin et al., 2010). As food and energy prices are more volatile than the overall inflation rate and often suffer from large hikes, respondents would more often use the outer bins in their prognosis. Overall, we can reject the null hypothesis that the distribution of the used number of bins is the same across the two wordings at the 1%-level. Finally, in both subgroups, the proportion of respondents reporting 50-50% answers is relatively low at about 7.7% and 7.1%, respectively, and does not differ significantly across wordings.

2.5.5 Consistency across Quantitative Measures

Internal consistency across multiple inflation expectations measures is desirable for several reasons. From a survey designer's perspective, low internal consistency might indicate miscommunication. This could mean that the question does not accurately elicit the beliefs or expectations of interest, or in other words, has low *face validity.* Obviously, different interpretations on the side of the respondents, as well as their personal characteristics, could influence the degree of consistency found in the data (Zhao, 2022). Nonetheless, it is an important aspect to consider, especially when the data is used for policy evaluation and recommendations.

As a proxy for the inconsistency of the answers, I report the proportion of occurrences when the point prediction (i) does not lie in the range of the support (fitted or self-reported), or (ii) lies outside of the difference between the estimated p_5 and p_{95}. First, let's focus on subjective histograms. For the *inflation rate* treatment, the first share is at 11%, whereas for *prices in general* at 15%, which is somewhat surprising given it was expected that the latter will yield higher internal consistency, due to its simplicity. However, it might also be the case that respondents report more extreme values for their point prediction, thinking of food and gas prices, and are thus subsequently more prone to revising their expectations downwards when confronted with the bin response scale. That is, priming effects are higher for the *prices* wording. In contrast, in the min-max setup, this pattern is reversed: It is the *prices in general* wording that produces answers more in line with peoples' initial point forecasts and yields higher internal consistency of responses. In the spirit of Zhao (2022), I also report the number of occasions where there is zero probability assigned to the interval containing the point forecast and the number of cases of positive probability in disjoint regions. The latter do occur in the data, but rather seldom in about 3% of all observations in either wording. Also, roughly 16% of respondents assign zero probability to the interval that contains their point prediction, but again, there are no significant differences with respect to wording.

What is striking is the fact that irrespective of treatment, roughly 30% to 40% of the participants in each subgroup report point forecasts that are either smaller than q_5 or exceed q_{95}. In contrast to professional forecasters whose point predictions are known to represent some central measure of tendency such as the mean, many consumers seem to provide a point prediction that reflects a tail outcome. This is an important implication as point forecasts and derived means from distributions are often used as substitutes in the empirical and experimental literature. To this end, in Figure 2.6 in the

Table 2.2: Averages of support endpoints, moments, and percentiles

	mean	sd	left endpoints	right endpoints	$p5$	$p25$	$p50$	$p75$	$p95$
			probabilistic question						
A1: *inflation rate*	2.09	1.84	−3.31	8.49	−0.87	0.74	2.06	3.41	5.14
C1: *prices in general*	3.50	2.07	−2.25	11.13	0.22	1.96	3.44	4.99	6.97
			min-max question						
B1: *inflation rate*	4.31	0.87	2.20	6.28	2.85	3.67	4.33	4.96	5.71
D1: *prices in general*	5.51	1.34	2.24	8.54	3.25	4.53	5.54	6.52	7.67

Appendix, I plot the point predictions against the means derived from the probabilistic and min-max questions and the corresponding correlations. With 68%, the combination of the min-max format with the *prices in general* wording yields the highest correlation among the four options. One possible advantage of the min-max over the probabilistic question is that people directly state the range of the support making answers across multiple questions more coherent. This is an essential point as observations that do not abide by this internal consistency rule are often filtered out of the sample. For the case of the min-max question, the proportion of data that is thrown out would be (almost) cut by half depending on the wording, compared to the probabilistic setup.

2.5.6 Distributions of Derived Measures

Table 2.2 reports the average values for the distributions of fitted moments, endpoints, and percentiles across all four treatment arms. It seems that irrespective of the question format, the *prices in general* wording produces (i) higher expected values for the inflation rate, but (ii) lower uncertainty in terms of the standard deviation. The distributions of the fitted means are additionally plotted in Figure 2.7 in the Appendix. In all four cases, across wording and format variations I can reject the null hypothesis of a two-sample Kolmogorov-Smirnov Test that the distributions are equivalent at the 1%-level. Nonetheless, caution is advised when interpreting these results, since they are very sensitive to the underlying assumptions on the support of the fitted distribution for the case of the probabilistic question. Another possibility in the quantification procedure is to use the highest and lowest reported point forecasts as endpoints for the histograms. While feasible, it is not clear that this approach is more desirable than the one used in the current paper or others, for that matter, as assumptions on the support will change with each survey wave and diminish comparability over time. This again speaks to the fact that the min-max question has an advantage over the probabilistic setup, namely, the lack of necessity for assumptions on the support of the histogram, thus ensuring comparability over time and across studies.

Furthermore, I observe that both the left endpoint and the p_5 are (often) negative on average for the probabilistic setup, whereas for the min-max question, both of them are positive and even slightly above 2%. While there is little difference in the p_{95}, the one in the right endpoint is striking, it is about $2-3$ pp larger in the probabilistic format than in the min-max setup. This adds further evidence to the notion that providing respondents with pre-defined bins might have a priming effect on their responses and a format free of assumptions on the support might be more desirable.

2.6 Effects of Wording and Format on Consumers' Inflation Expectations

2.6.1 ATEs in Probabilistic Expectations

Based on Equation (1), I run a simple linear regression to estimate the average treatment effects (ATEs) of changes in wording and format for different measures derived from probabilistic responses. Table 2.3 reports the estimated coefficients for mean expected inflation derived as described in Section 2.3. Column (1) of Table 2.3 uses the raw, unweighted data such that the estimates correspond to those in column (1) of Table 2.2

The baseline category stands for the treatment group A1 which received the combination of the probabilistic question and the *inflation rate* wording that is currently the default version in leading consumer surveys such as the SCE, CSCE, and BOP-HH. On average, respondents from this treatment arm expect an inflation rate of slightly more than 2% or 1.8% when accounting for influential observations. However, adopting the min-max format more than doubles people's mean expected inflation. Substituting current wording with the *prices in general* one adds an additional $1.0 - 1.4$ pp to the already elevated

Table 2.3: ATEs on derived mean expected inflation

	Dependent variable: $E[\pi_{t+12}]$		
	OLS	*OLS weighted*	*robust linear*
	(1)	(2)	(3)
intercept	2.088^{***}	2.193^{***}	1.838^{***}
	(0.216)	(0.219)	(0.074)
format	2.222^{***}	2.459^{***}	1.076^{***}
	(0.311)	(0.315)	(0.107)
wording	1.412^{***}	1.285^{***}	0.984^{***}
	(0.305)	(0.310)	(0.105)
joint effect	3.425^{***}	4.239^{***}	2.334^{***}
	(0.308)	(0.310)	(0.105)
Observations	3,686	3,686	3,686
R^2	0.034	0.052	
Adjusted R^2	0.034	0.051	
Residual Std. Error (df = 3682)	6.634	6.639	1.855

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors in parenthesis.

expectations. The joint effect is highly significant meaning the expectations of the final treatment group, where I vary both wording and format exceed the one of the default by about 2.3 – 4.2 pp. However, it rarely exceeds the sum of the other two coefficients on format and wording. When testing formally whether combing the two treatments in group D1 has a larger effect than implementing them separately, one cannot reject the null hypothesis, except in specification (3), where the lack of difference can be rejected at the 10%-level. This means that there is little evidence for a sizeable interaction effect beyond the separate estimates for format and wording. While the estimated coefficients are prone to changes in magnitude as soon as I control for the presence of outliers using Huber weighted regression in the specification (3), they remain statistically significant and large in absolute value.

While the changes in wording and format always lead to an increase in the predicted inflation rate on average, that is not the case for uncertainty measured by the standard deviation of the fitted distribution. The introduction of the *prices in general* wording leads to significantly higher individual uncertainty. As discussed in earlier research (Bruine de Bruin et al., 2010), people may interpret this question as asking them more about their personal shopping experiences. Therefore, the sign of the coefficient on wording is intuitive given that food and energy prices usually display higher volatility, as well as the fact that different respondents may have different consumption baskets. Whereas a change in wording increases uncertainty, switching to a min-max format reduces it by 0.6 to 1 pp, depending on the regression specification reported in Table 2.4. Regarding the interaction of the two in treatment group D1, the influence of the question format seems to prevail and the overall estimated effect is negative. In column (2), one can reject the null hypothesis that the interaction term equals the sum of the remaining two treatment effects at the 10%-level. Again, controlling for outliers in column (3) of Table 2.4 reduces the absolute value of the coefficients but does not change their sign or significance.

To sum up, it appears that reporting behavior is very prone to changes induced by variations in wording and format. The results suggest that mean expectations and uncertainty, both important indicators for policy-makers, vary strongly with survey framing. In some instances, changes in framing can lead to an increase in mean expectations of more than 100% compared to the initial level. Regarding the hypothesis stated in Section 2.3, one can confirm the initial assumption that simpler wording and less restrictive format would lead to higher expected inflation on average. Contrary to what was expected, respondents in the min-max treatment group appear much less uncertain on average based on the standard deviation of their histograms. This suggests that responses produced by the standard probabilistic question might be artificially more spread out, for instance, due to (i) assumptions on the support of the histogram, (ii) strong framing effects of the probabilistic format, or a combination of both.

Another aim of the RCT is to test whether treatment effects are symmetric across socio-demographic groups. As previously discussed in Section 2.3, ex-ante, one can expect that this is not the case, since answering a probabilistic question about the *inflation rate* requires a certain degree of numeracy and sophistication. Populations with lower financial literacy generally have more difficulty expressing their expectations

Table 2.4: ATEs on uncertainty

	Dependent variable: σ_π		
	OLS	*OLS weighted*	*robust linear*
	(1)	(2)	(3)
intercept	1.837***	2.011***	1.286***
	(0.073)	(0.074)	(0.027)
format	−0.965***	−1.018***	−0.620***
	(0.106)	(0.107)	(0.039)
wording	0.237**	0.387***	0.406***
	(0.104)	(0.105)	(0.038)
joint effect	−0.493***	−0.373***	−0.230***
	(0.105)	(0.105)	(0.038)
Observations	3,686	3,686	3,686
R^2	0.040	0.048	
Adjusted R^2	0.039	0.048	
Residual Std. Error (df = 3682)	2.255	2.256	0.709

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors in parenthesis.

in terms of probabilities (Fischhoff and Bruine de Bruin, 1999; Bruine de Bruin et al., 2000). Those less knowledgeable of economic terms such as the *inflation rate* might be thus more susceptible to framing effects. Unfortunately, while the core module of the survey collects information on multiple respondents' characteristics such as age, income, employment, profession, etc., there is no explicit measure of one's financial literacy. Instead, I will focus on the reported educational background and household income category as proxies. Additionally, as many studies document differences in expectations by gender, e.g. originating from different shopping habits (D'Acunto et al., 2021b; Bryan and Venkatu, 2001), I also assess the treatment effects for men and women separately. Finally, as literature reports differences in expectations based on experiences that have a long-lasting or 'imprinting' effect on some populations such as the German reunification (Goldfayn-Frank and Wohlfart, 2020) - I compare the effects for respondents who lived in East Germany before 1990 to those who did not. All estimates from the sample splits listed above for mean inflation and uncertainty are reported in Tables 2.8 and 2.9 in the Appendix, respectively. Notably, across all sub-samples, the effects of wording and format, as well as the estimated joint influence are highly significant and large in magnitude for $\mathrm{E}[\pi_{t+12}]$ but not always for $\sigma^{\pi_{t+12}}$.

Table 2.10 reports the differences between the estimated treatment effects for several pairs of groups - by education, income, gender, and by experience. Interestingly, one can observe different patterns depending on whether we look at mean expected inflation (top panel) or uncertainty (bottom panel) in Table 2.10. Asymmetry in the treatment effect of the question format is observed for mean expectations by gender and education, whereby the increasing effect is stronger for females and those with no college degree by roughly 2 pp, relative to their counterparts. The joint effect appears to increase the upward bias of mean expected inflation and it does so a lot more strongly (by 2.5 pp) for below-median income households than for the top-50% in the income distribution. There are no asymmetries observed for wording, except for those living in East Germany prior to 1989, for which the *prices* option increases uncertainty by more than 0.6 pp compared to the counterpart. While asymmetry in the treatment effect of wording appears much less pronounced than initially predicted, the estimates for format provide substantial evidence that different question format produces very different mean expectations depending on the underlying population. Females and those with no college degree appear to adapt their forecasts for inflation more closely to the underlying response scale in the probabilistic question.

2.6.2 Effect of Wording on Forming Expectations

Next, I move on to evaluate Question 2 and analyze whether differences observed between the two wording treatments can be accounted for by differences in interpretation. Of particular interest is, whether respondents associate the *prices in general* formulation most with price changes of specific goods they observe in their day-to-day shopping. The distributions of the topics respondents self-reportedly thought about when answering inflation expectation questions are depicted in Figure 2.3. Overall, the resulting frequencies of topics under the *prices in general* formulation appear similar to the one reported by Bruine de Bruin et al. (2010) using the same wording, with the topic 'inflation rate',

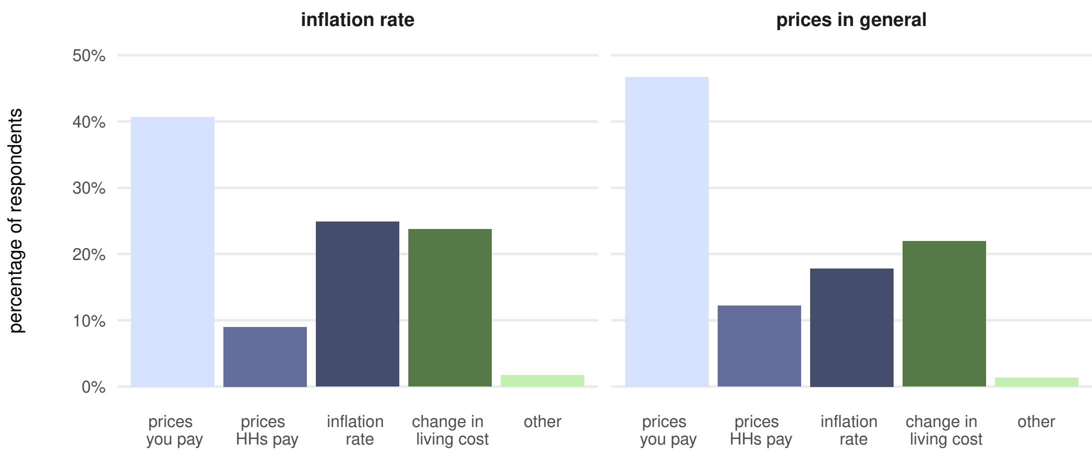


Figure 2.3: Effect of wording on forming expectations. Each respondent can select one of five options or provide an individual answer.

ranked third instead of fourth. The distribution observed in the German data seems slightly more polarized with the majority of respondents thinking about their personal shopping experience, compared to 'Prices Americans pay', which was selected by roughly 40% of US survey participants as the main topic[14].

Although for the *inflation rate* wording, the corresponding share is somewhat smaller (41% vs 47%), overall a substantial percentage of participants think about supermarket or gas prices when producing their inflation forecast. Even when asked directly for their inflation rate prediction, only about one-quarter of the respondents report actually thinking about it. Considering the additional information people receive at the beginning of the survey about what inflation is and that the questions are continuously accompanied by info boxes, this share is surprisingly small. While the *inflation rate* wording is indeed able to reduce the share of respondents producing a forecast based solely on their personal shopping experience, broader concepts of price changes remain less thought of.

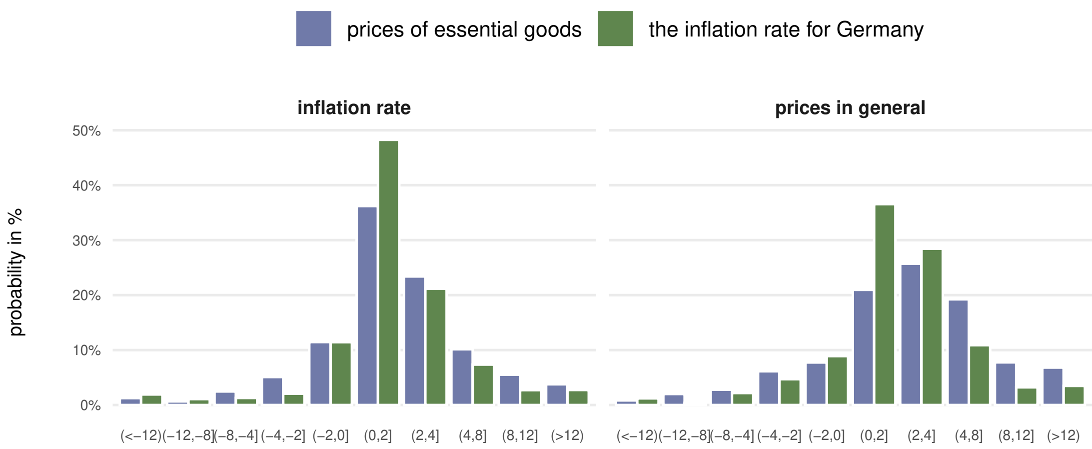
More formally, to estimate whether different wording choices impact the question interpretation, I model the probabilities of selecting the discrete, non-ordered options

[14]Note that the question in Bruine de Bruin et al. (2010) is framed differently than the one conducted in BOP-HH. In their setup, one could rank multiple topics, whereas, in the current setting, one could only select one topic. However, due to concerns about survey fatigue, the question implemented in BOP-HH had to be simplified.

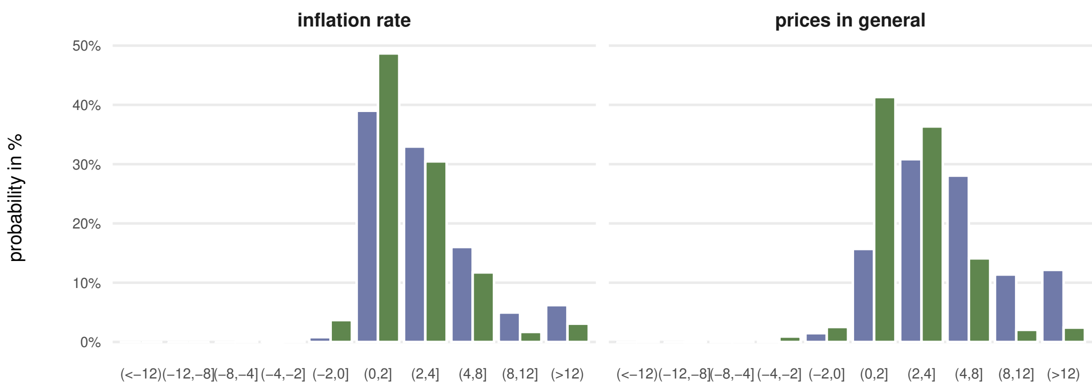
listed in Question 2 in a multinomial-logit setting. The log-odds estimates are reported in Table 2.11 in the Appendix. The base level of the variable from Question 2 is set to be the 'inflation rate'. That is, all log odds are reported relative to the probability of selecting this option. The results are consistent with the graphical evidence. The log odds of selecting the option 'prices you pay' are 0.60 for the default wording and increase by another 0.52 when we apply the *prices in general* one. As expected, the log odds for the remaining options 2 and 4 are positive as well, meaning that the corresponding wording indeed causes significant variation in the interpretation. Taking together both the graphical and quantitative evidence, the share of people reporting they associate the question with the actual inflation rate remains remarkably low. The results suggest that (i) diffuse question interpretation continues to be a major source of variation, beyond true differences in beliefs, and (ii) people base their inflation forecast on specific prices they observe in their day-to-day shopping instead of broader concepts about price changes.

In order to better quantify the effect mentioned in (i) above and highlight the importance of clear and precise wording, in Figure 2.4 I plot the average subjective probabilities for people who reported thinking of 'prices of essential goods' vs 'the inflation rate for Germany' for the four treatment arms. Even though the respondents within each treatment group received the exact same questions, the resulting pairs of distributions are very distinct. More formally, using a HotellingsT2 test one can reject the null that the pairs of subjective probabilities are the same within each wording and format combination. Expectations of people thinking about the inflation rate are clearly more anchored and contain less probability mass in the right tail than those reporting thinking about essential goods' prices across all treatments.

Table 2.12 in the Appendix documents summary statistics for mean expected inflation and uncertainty for different groups of respondents, based on which concepts they reported thinking about most. Interestingly, in most cases, there appear to be differences by wording even if participants report thinking about the same concept. For example, in the probabilistic question (min-max question), those who base their forecast primarily on observed essential goods' prices report higher expected inflation in the *prices in general* wording 3.4% (6.5%) than in the *inflation rate* wording 2.2% (4.6%) on average. The same is also true for individual uncertainty. However, the influence of wording choice



(a) probabilistic format (default)



(b) min-max format

Figure 2.4: Average subjective probabilities by wording and association. Left panels represent groups who received the *inflation rate* wording, right panels - *prices in general* wording. Blue bars represent the subjective probabilities reported by participants who thought most about the prices of essential goods such as food and gasoline. Green bars depict the probabilities for those who report thinking about the inflation rate in Germany.

is not present for the group reporting thinking about the inflation rate. This could be interpreted in a sense that the broader concept of 'inflation rate' not only acts as a numerical anchor but can potentially reduce more general framing effects.

Finally, regarding differences across demographic groups, e.g. by education, we observe the following: for each wording and each education group, the leading choice is 'prices you pay' as shown in Figure 2.8 in the Appendix. However, the proportion of respondents selecting this option declines as education increases. Also, people with lower education tend to think less often about the actual inflation rate, especially in the *prices in general wording.* Again, this suggests substantial asymmetry in the effects of wording for different demographic groups.

2.6.3 Point Forecasts and Perceptions

One can further compare the distributions of point forecasts which result from providing respondents with different wordings. These are depicted in Figure 2.5. Summary statistics can additionally be found in Table 2.13 in the Appendix. The findings are in line with those from existing literature. The distribution of point forecasts from the *inflation rate* wording in the left panel of Figure 2.5 is much more concentrated on values between 0 and 4%, with more than 40% of respondents reporting values in the interval containing the inflation target. For *prices in general*, the majority of reported point forecasts lie in the bins covering values between 0 and 6%, and they appear roughly equally distributed across the three intervals in the range. While higher inflation values seem more likely for the *prices in general* treatment, there are more participants expecting deflation in the counterpart treatment. A spike in reported point predictions occurs for bins containing values that are multiples of five in both distributions, phenomena indicating uncertainty as documented by Binder (2017), is present in both distributions. Overall, one can reject the null hypothesis that the distributions are the same at the 1%-level using a Mann-Whitney test. Notably, the difference of more than 2 pp in mean point forecasts across wordings appears to be much more pronounced than those documented in previous literature, e.g. Bruine de Bruin et al. (2017) using representative data from the Netherlands.

One important policy indicator when it comes to aggregate distributions is the *disagreement* of point forecasts, that is the standard deviation across the latter. Disagreement is found to be a good proxy for the uncertainty of professional forecasters, especially in times of economic distress (Boero et al., 2011), and is reasonably well correlated with

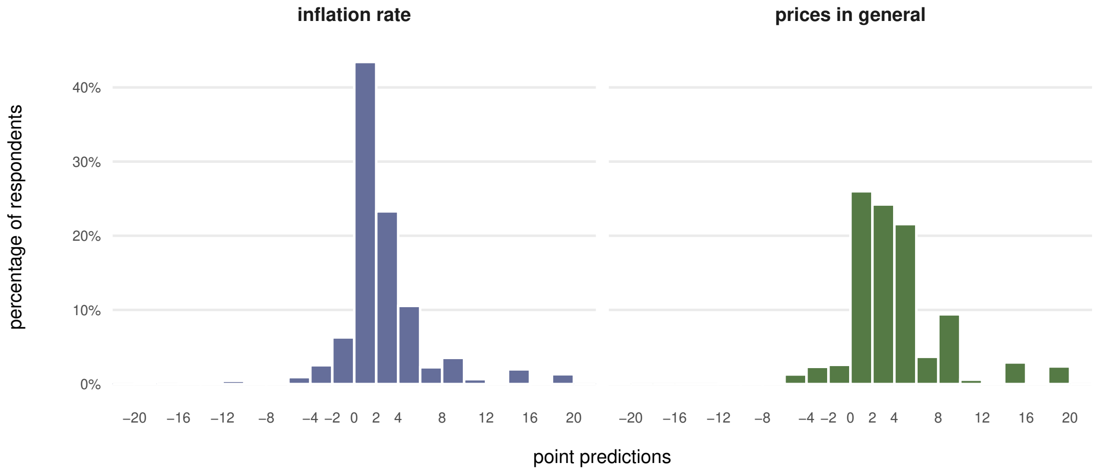


Figure 2.5: Distribution of point forecasts across wordings. Point predictions exceeding 100% in absolute value are removed. Values above 20% and below −20% are not shown but accounted for in the graph (roughly 2% of all observations). Observations are not weighted.

rounding-based uncertainty measures in the context of consumers (Binder, 2017). In each scenario, the *prices in general* wording yields point forecasts that are significantly more dispersed - 7.08 vs 6.66 pp, confirmed by a Fligner-Killeen Test. In Figure 2.9 in the Appendix I provide a more detailed overview of the point forecasts' distributions by both wording and interpretation, the corresponding measures are reported in Table 2.13. To illustrate the importance of question interpretation as a source of heterogeneity, again I compare the distributions of individuals who report thinking about the same topic across wordings. It seems that the *prices* wording produces forecasts that (i) exhibit a higher upward bias, consistent with previous literature (Dräger and Fritsche, 2013), and (ii) higher disagreement, however only when people focus on personal shopping experiences. Another takeaway from Table 2.13, is that similar to derived means, the effect of wording seems to disappear when survey respondents actually focus on the inflation rate when forming their point forecasts. This reiterates the importance of aiming for wording that increases the share of respondents thinking about the inflation rate. The bottom panel of Table 2.13 documents the same measures for the inflation perceptions of respondents, for which, to a large extent, the analogous patterns apply.

2.7 Conclusion

Directly eliciting consumer expectations about inflation has evolved tremendously in the past few decades. The adoption of a probabilistic format in multiple large-scale consumer surveys has furthered inflation expectations research, deepened our understanding of the underlying formation process, and improved cross-country comparison. Still, several aspects in this context need to be addressed. For once, there appears to be a large priming effect on responses due to the underlying bin definitions which in some cases can act as an anchor to consumers' expectations. Additionally, the assumptions on both the expectation formation process as well as the subsequent processing of the data often appear strong. In particular, the former relies on survey participants holding precise probabilities about future events and being able to convey them in a numerical format.

Even though it yields a higher non-response rate, the min-max question provides a viable alternative to the current default format. As pointed out by Hayo and Méon (2022), a guided survey question might potentially introduce more noise to the data instead of collecting informative answers. The same argument may apply to the probabilistic format. The min-max setup is attractive for survey designers due to its simplicity and straightforwardness. The format reduces the priming effects that result from the underlying scale to a minimum as well as eliminates 'problematic' cases such as bi-modal distributions or those with positive probability in disjoint regions. Another important aspect is that the probabilistic format strongly relies on a stable, tractable set of bin definitions over time. This can prove challenging as inflation increases as is currently the case in the euro area and in the US. While a change in bin definitions could cause a break in the time series, as people assign more probability to outer intervals, the assumptions on their bounds become more important. As inflation varies different choices for these bounds could be justified, diminishing comparability over time.

While a large-scale implementation of an alternative question to elicit subjective distributions of future inflation will be costly, using the min-max format as an occasional 'sanity check' for expectations could be advantageous, especially when the data is used for policy evaluation or recommendations. The results on question wording confirm findings of previous experimental studies and reiterate the importance of precise question formulation. Even a direct formulation such as the *inflation rate* seems insufficient to invoke thoughts about inflation among the majority of consumers. More so, roughly 40% of the survey participants still mainly think about specific prices such as those of food and gas when producing their forecast. A potential avenue for future research could be to explore how to reduce this share via wording or providing respondents with additional information.

2.8 Appendix

2.8.1 Additional Graphs

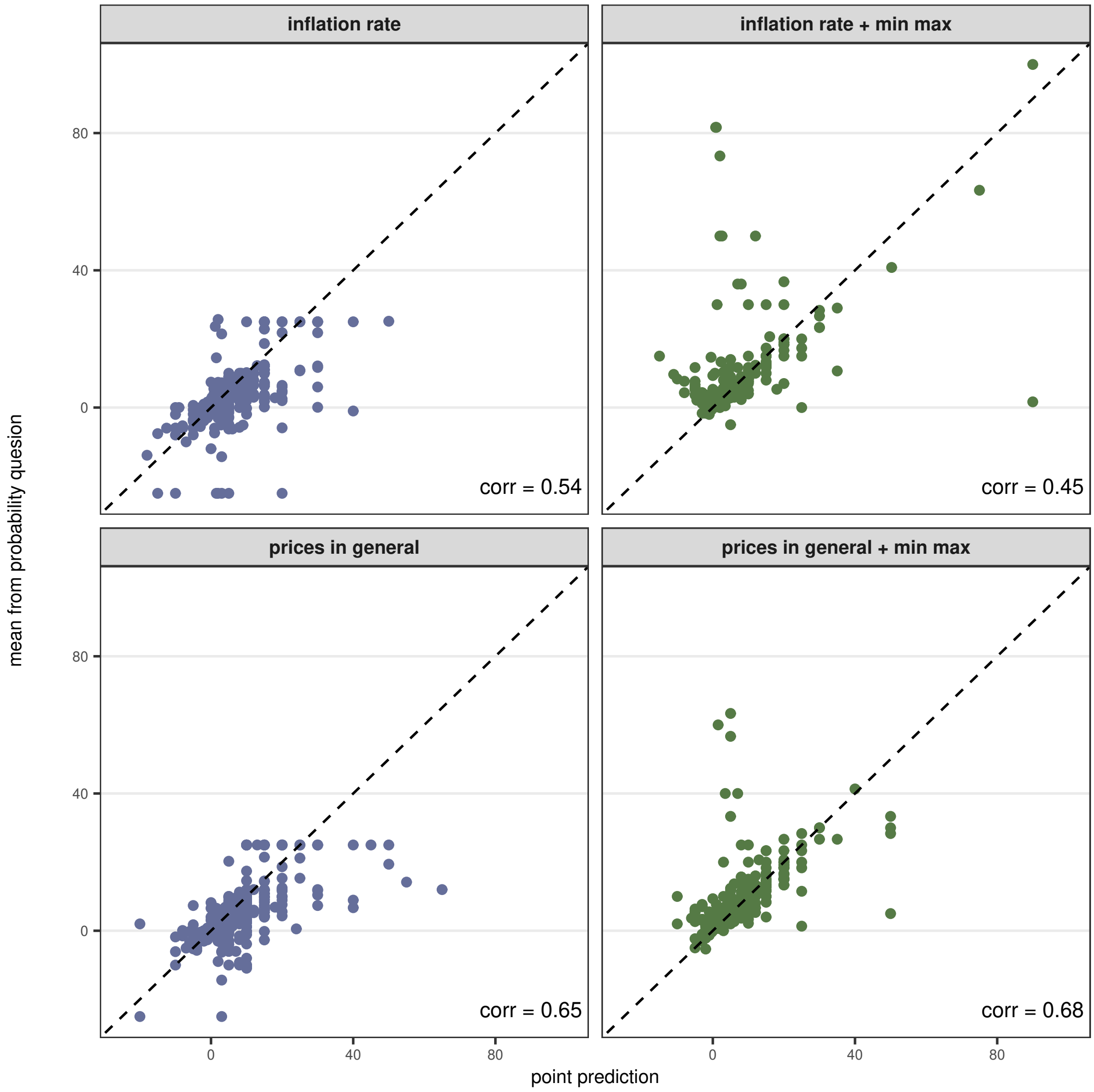


Figure 2.6: Point predictions and means from derived or self-reported distributions together with correlation coefficients.

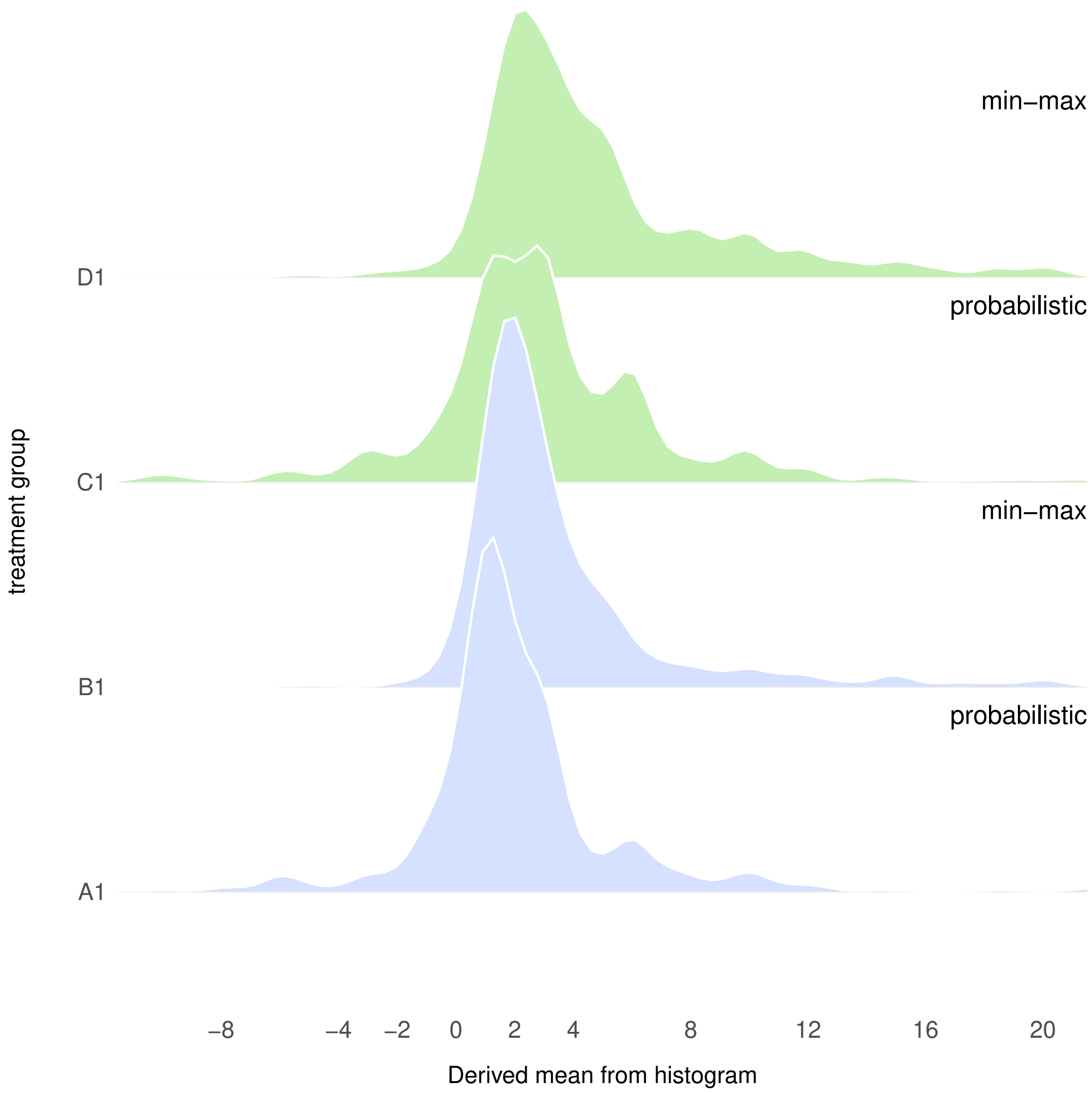
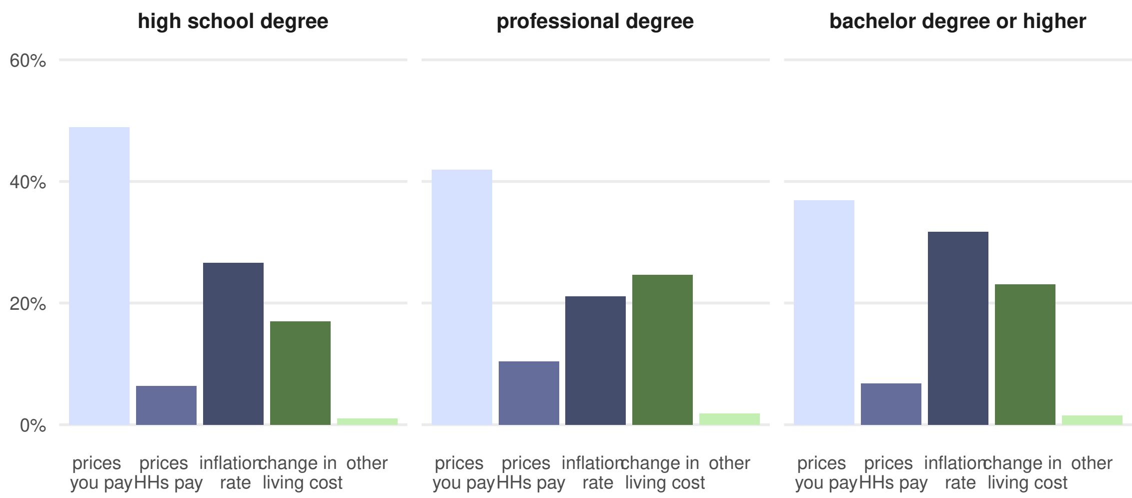
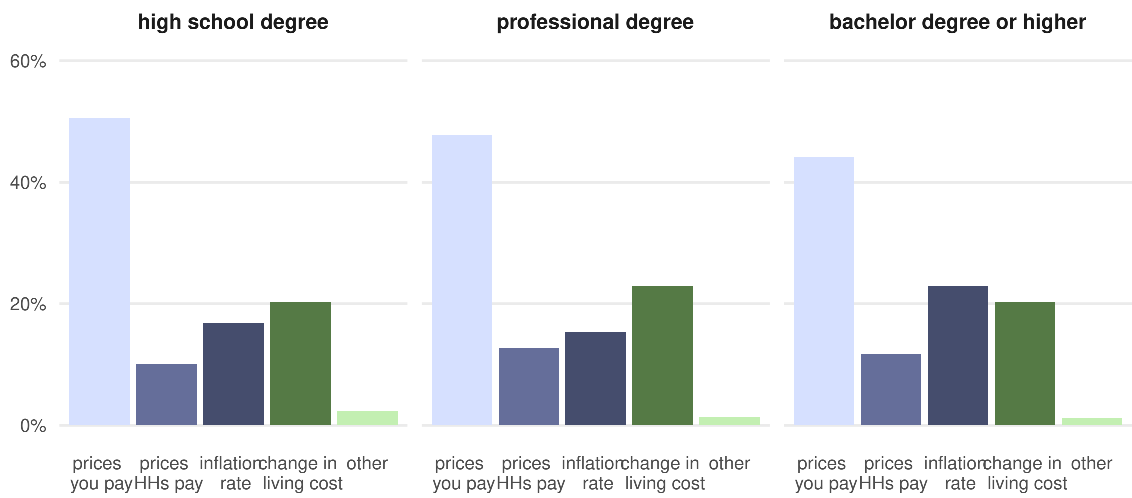


Figure 2.7: Distribution of derived means from histograms. Treatment groups A1 and B1 receive the *inflation rate* wording (depicted in blue), and the remaining groups, C1 and D1, receive the *prices in general* (depicted in green).



(a) inflation rate



(b) prices in general

Figure 2.8: Effect of wording on forming expectations for different demographic groups. Each respondent can select one of four pre-defined options or provide an individual answer. Upper panel: treatment groups A1 and B1, which received the *inflation rate* wording. Lower panel: groups C1 and D1 who answered a *prices-in-general* question.

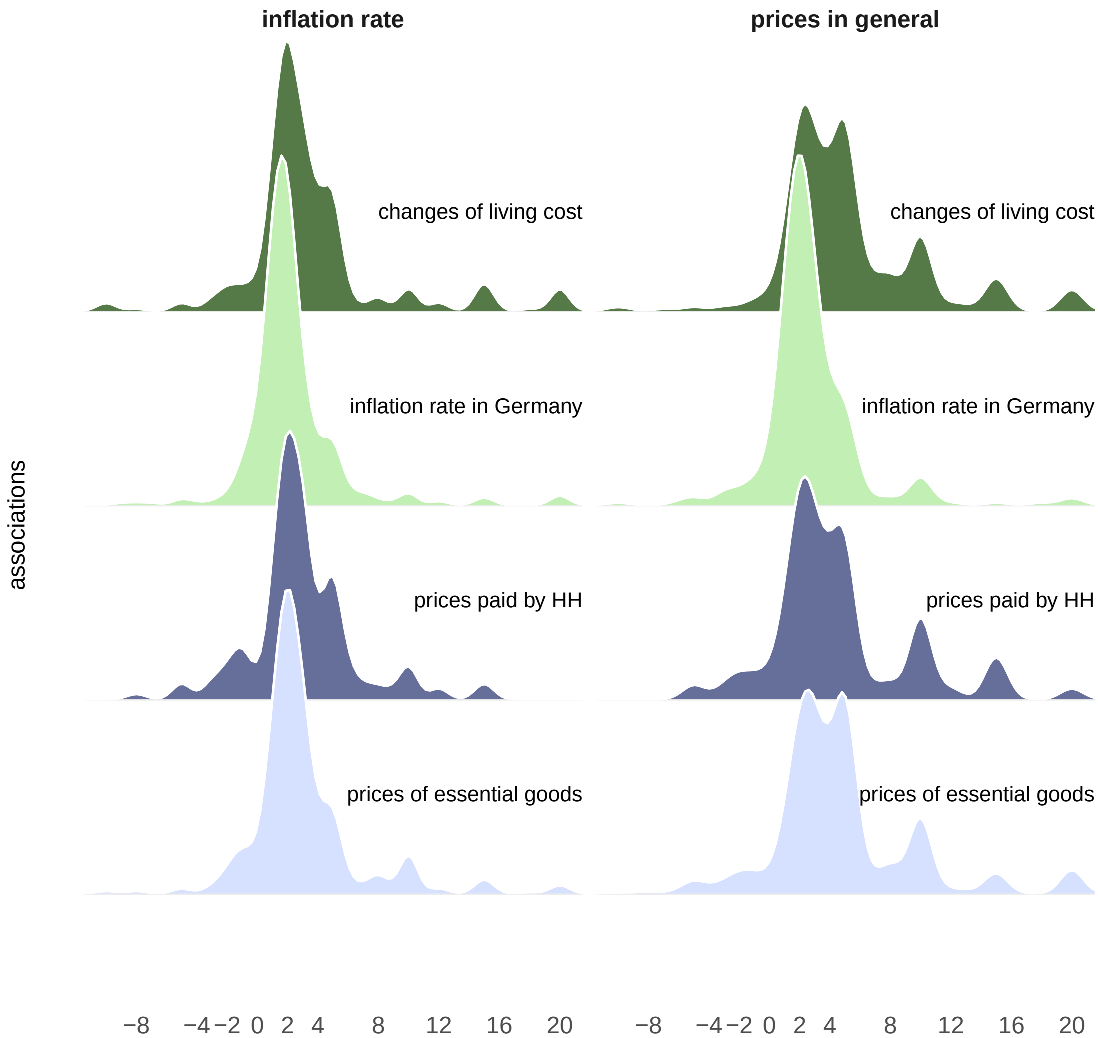


Figure 2.9: Distributions of point forecasts by wording and association. The left panels depict distributions from groups that receive the *inflation rate* wording, right panels - *prices in general*.

2.8.2 Additional Tables

Table 2.5: Average marginal effects of demographics on item non-response

	Dependent variable: $d_{non-response}$			
	inflation rate		prices in general	
	probabilistic	min-max	probabilistic	min-max
	(1)	(2)	(3)	(4)
age (in years)	0.003***	0.002**	0.001	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
professional degree	−0.016	−0.039	−0.066	0.097*
	(0.040)	(0.055)	(0.046)	(0.057)
bachelor or higher	−0.048	−0.031	−0.082***	0.040
	(0.031)	(0.051)	(0.028)	(0.089)
employed	−0.012	−0.010	−0.039**	−0.020
	(0.021)	(0.028)	(0.019)	(0.025)
male	−0.035**	−0.067***	−0.042***	−0.035*
	(0.016)	(0.022)	(0.016)	(0.019)
HH income €2,500 to €3,500	−0.044***	−0.050**	−0.053***	−0.010
	(0.016)	(0.025)	(0.015)	(0.023)
HH income €3,500 to €5,000	−0.068***	−0.039	−0.048***	−0.010
	(0.015)	(0.027)	(0.016)	(0.025)
HH income €5,000 or more	−0.066***	−0.076***	−0.068***	−0.018
	(0.015)	(0.027)	(0.013)	(0.029)
HH size	0.019**	0.013	0.003	0.024**
	(0.009)	(0.012)	(0.009)	(0.010)
West	−0.023	−0.055**	0.011	0.001
	(0.022)	(0.027)	(0.025)	(0.028)
South	0.016	−0.016	−0.0001	−0.004
	(0.024)	(0.028)	(0.022)	(0.026)
East	0.032	−0.035	0.019	0.057
	(0.030)	(0.030)	(0.029)	(0.037)
born in East Germany pre-1989	0.072	0.018	0.069	−0.096***
	(0.084)	(0.124)	(0.104)	(0.009)
homeowner	−0.007	−0.019	0.024	−0.025
	(0.018)	(0.025)	(0.016)	(0.023)
Observations	1,001	999	1,003	997
Log Likelihood	−242.022	−387.675	−251.049	−317.829
Akaike Inf. Crit.	514.045	805.350	532.099	665.658

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors in parenthesis.

Table 2.6: Rounding in the min-max setup

	Dependent variable: IQR		
	OLS	*OLS weighted*	*robust linear*
	(1)	(2)	(3)
intercept	0.167	−0.044	0.589***
	(0.531)	(0.472)	(0.074)
round	1.195***	1.265***	0.564***
	(0.091)	(0.096)	(0.015)
point forecast	0.089***	0.102***	0.117***
	(0.011)	(0.010)	(0.002)
m5	0.382*	0.354	−0.006
	(0.205)	(0.217)	(0.034)
prices in general wording	0.083	0.207	0.140***
	(0.136)	(0.150)	(0.024)
age (in years)	−0.001	−0.001	−0.005***
	(0.006)	(0.006)	(0.001)
professional degree	−0.180	−0.111	−0.117**
	(0.347)	(0.290)	(0.046)
bachelor or higher	−0.280	−0.347	−0.106**
	(0.355)	(0.304)	(0.048)
employed	0.051	0.0001	0.031
	(0.175)	(0.182)	(0.029)
male	−0.058	0.028	−0.007
	(0.138)	(0.149)	(0.023)
HH income €2,500 to €3,500	0.114	0.132	−0.059*
	(0.192)	(0.213)	(0.033)
HH income €3,500 to €5,000	0.130	0.219	−0.013
	(0.200)	(0.219)	(0.034)
HH income €5,000 or more	0.101	0.126	−0.046
	(0.233)	(0.260)	(0.041)
HH size	−0.032	0.018	0.033***
	(0.072)	(0.076)	(0.012)
West	0.279	0.445**	−0.012
	(0.199)	(0.221)	(0.035)
South	0.025	0.083	0.003
	(0.189)	(0.212)	(0.033)
East	0.189	0.216	0.039
	(0.276)	(0.287)	(0.045)
born in East Germany pre-1989	−0.010	−0.101	0.047
	(0.257)	(0.281)	(0.044)
homeowner	−0.108	−0.282*	−0.001
	(0.159)	(0.167)	(0.026)
Observations	1,712	1,712	1,712
R^2	0.236	0.267	
Adjusted R^2	0.228	0.259	
Residual Std. Error (df = 1693)	2.734	2.987	0.380

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$. Standard errors in parenthesis.

Table 2.7: Effect of wording on number of used bins

	Dependent variable: number of used bins	
	Poisson	*quasi-Poisson*
	(1)	(2)
intercept	1.876***	1.497***
	(0.067)	(0.075)
prices in general wording	0.122***	0.132***
	(0.025)	(0.028)
age (in years)	−0.015***	−0.012***
	(0.001)	(0.001)
professional degree	−0.056	0.048
	(0.043)	(0.052)
bachelor or higher	−0.069	0.122**
	(0.045)	(0.053)
employed	−0.016	−0.063*
	(0.029)	(0.032)
male	0.038	0.046
	(0.025)	(0.028)
HH income €2,500 to €3,500	0.053	0.061
	(0.036)	(0.039)
HH income €3,500 to €5,000	0.094**	0.027
	(0.038)	(0.042)
HH income €5,000 or more	0.093**	0.053
	(0.042)	(0.047)
HH size	−0.015	0.037***
	(0.013)	(0.014)
West	0.095**	0.033
	(0.041)	(0.045)
South	0.036	0.024
	(0.039)	(0.043)
East	−0.013	0.003
	(0.053)	(0.057)
born in East Germany pre-1989	−0.016	−0.015
	(0.048)	(0.051)
homeowner	0.039	0.003
	(0.028)	(0.031)
Observations	1.831	1.753
Log Likelihood	-3,892.672	-3,259.051
Akaike Inf. Crit.	7,817.345	6,550.103

Note:*p<0.1; **p<0.05; ***p<0.01. Standard errors in parenthesis.

Table 2.8: Demographic splits for mean expected inflation

	Dependent variable: $E[\pi_{t+12}]$			
	By gender		By education degree	
	male	female	no college	college
	(1)	(2)	(3)	(4)
intercept	2.620***	0.430	2.409***	1.529*
	(0.721)	(1.100)	(0.829)	(0.908)
format	1.819***	3.527***	3.079***	1.077**
	(0.353)	(0.563)	(0.411)	(0.451)
wording	1.189***	1.238**	1.421***	1.165***
	(0.354)	(0.546)	(0.402)	(0.447)
joint effect	3.967***	4.425***	4.728***	2.737***
	(0.352)	(0.549)	(0.398)	(0.455)
Observations	2,088	1,461	2,372	1,156
R^2	0.109	0.112	0.108	0.067
Adjusted R^2	0.102	0.102	0.102	0.055
	By income		Born pre-1989 in	
	lower 50%	upper 50%	east	west
intercept	1.489	2.643***	4.870	2.672***
	(0.991)	(0.800)	(3.910)	(0.625)
format	3.180***	1.954***	3.499***	2.431***
	(0.499)	(0.352)	(1.045)	(0.318)
wording	1.370***	1.021***	3.068***	0.824***
	(0.491)	(0.349)	(1.026)	(0.313)
joint effect	5.258***	2.728***	3.674***	4.187***
	(0.481)	(0.359)	(1.040)	(0.314)
Observations	1,888	1,661	609	2,940
R^2	0.099	0.111	0.120	0.106
Adjusted R^2	0.092	0.104	0.097	0.101

Note:*p<0.1; **p<0.05; ***p<0.01. Standard errors are reported in parenthesis. Observations are weighted using survey weights. Socio-demographic controls are included.

Table 2.9: Demographic splits for uncertainty

	Dependent variable: $\sigma^{\pi_{t+12}}$			
	By gender		By education degree	
	male	female	no college	college
	(1)	(2)	(3)	(4)
intercept	2.589***	2.994***	3.241***	2.257***
	(0.264)	(0.361)	(0.295)	(0.238)
format	−0.938***	−1.116***	−1.034***	−1.084***
	(0.129)	(0.185)	(0.146)	(0.118)
wording	0.529***	0.084	0.358**	0.209*
	(0.130)	(0.179)	(0.143)	(0.117)
joint effect	−0.085	−0.785***	−0.322**	−0.655***
	(0.129)	(0.180)	(0.142)	(0.119)
Observations	2,088	1,461	2,372	1,156
R^2	0.088	0.066	0.068	0.150
Adjusted R^2	0.081	0.056	0.062	0.139
	By income		Born pre-1989 in	
	lower 50%	upper 50%	east	west
intercept	2.745***	3.305***	3.399***	3.105***
	(0.340)	(0.270)	(0.785)	(0.240)
format	−0.953***	−1.061***	−0.703***	−1.063***
	(0.171)	(0.119)	(0.210)	(0.123)
wording	0.359**	0.330***	0.878***	0.252**
	(0.168)	(0.118)	(0.206)	(0.120)
joint effect	−0.152	−0.806***	−0.320	−0.427***
	(0.165)	(0.121)	(0.209)	(0.121)
Observations	1,888	1,661	609	2,940
R^2	0.055	0.166	0.195	0.069
Adjusted R^2	0.048	0.159	0.173	0.064

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors are reported in parenthesis. Observations are weighted using survey weights. Socio-demographic controls are included.

Table 2.10: Heterogeneity in ATEs

	Dependent variable: $\mathrm{E}[\pi_{\mathrm{t+12}}]$		
	(1) format	(2) wording	(3) joint
by gender	-1.71^{**}	-0.05	-0.46
	(0.66)	(0.65)	(0.65)
by education	2.00^{***}	0.26	1.99^{***}
	(0.61)	(0.60)	(0.60))
by income	1.23	0.35	2.53^{***}
	(0.61)	(0.60)	(0.60)
by imprinting experience	1.07	2.24	-0.51
	(1.09)	(1.07)	(1.09)
	Dependent variable: $\sigma^{\pi_{t+12}}$		
	(1) format	(2) wording	(3) joint
by gender	0.18	0.45	0.70^{**}
	(0.23)	(0.22)	(0.22)
by education	-0.05	-0.15	-0.33
	(0.19)	(0.19)	(0.19)
by income	-0.11	-0.03	-0.65^{**}
	(0.21)	(0.21)	(0.20)
by imprinting experience	0.36	0.63^{**}	0.11
	(0.24)	(0.24)	(0.24)

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors of the differences between coefficients are reported in parenthesis.

Table 2.11: Multinomial logit regression

	Dependent variable: associations			
	1 prices you pay	2 prices paid by HHs	4 cost of living	5 other prices
intercept	0.603***	−0.931***	0.032	−2.929***
	(0.061)	(0.093)	(0.069)	(0.219)
prices in general	0.523***	0.682***	0.267**	0.226
	(0.091)	(0.128)	(0.104)	(0.320)
Akaike Inf. Crit.	9,405.255	9,405.255	9,405.255	9,405.255

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors in parenthesis.

Table 2.12: Summary statistics on derived means and std. deviation

	$E[\pi_{t+12}]$		$\sigma^{\pi_{t+12}}$	
	probabilistic format			
	inflation rate	prices in general	inflation rate	prices in general
prices of essential goods	2.20	3.41^{p1}	1.79	2.18^{p1}
prices paid by HHs in Germany	1.82	3.73^{p5}	1.89	2.01
inflation rate	1.38	2.15	1.88	1.60
changes in living cost	2.77	4.55^{p1}	1.90	2.20
	$E[\pi_{t+12}]$		$\sigma^{\pi_{t+12}}$	
	min-max format			
	inflation rate	prices in general	inflation rate	prices in general
prices of essential goods	4.60	6.49^{p1}	0.89	1.56^{p1}
prices paid by HHs in Germany	4.71	4.90	0.87	1.21^{p10}
inflation rate	3.55	2.92	0.76	0.99
changes in living cost	4.45	6.10^{p5}	0.96	1.30^{p5}

Note: ^{p1}, ^{p5}, ^{p10} indicate that the corresponding measure is significantly different in the *prices in general* from the *inflation rate* wording at the 1, 5, and 10%-level. To test for significant differences in means I use t-tests.

Table 2.13: Summary statistics of point forecasts and perceptions

	π^{pp}_{t+12}		$\sigma^{\pi^{pp}_{t+12}}$	
	inflation rate	prices in general	inflation rate	prices in general
prices of essential goods	3.40	5.61^{p1}	5.69	7.57^{p1}
prices paid by HHs in Germany	3.23	4.76^{p5}	7.82	4.78^{p10}
inflation rate	2.56	3.07	6.57	6.22
changes in living cost	4.19	6.09^{p1}	7.78	7.40^{p1}
all	3.36	5.15^{p1}	6.66	7.08^{p1}
	π^{pp}_{t-12}		$\sigma^{\pi^{pp}_{t-12}}$	
	inflation rate	prices in general	inflation rate	prices in general
prices of essential goods	3.64	5.64^{p1}	6.86	7.05^{p1}
prices paid by HHs in Germany	4.37	4.87	8.82	5.58^{p1}
inflation rate	3.02	3.05	7.76	3.95^{p1}
changes in living cost	3.92	6.15^{p1}	5.28	7.20^{p1}
all	3.60	5.20^{p1}	6.93	6.58^{p1}

Note:^{p1}, ^{p5}, ^{p10} indicate that the corresponding measure is significantly different in the *prices in general* from the *inflation rate* wording at the 1, 5, and 10%-level. To test for significant differences in means I use *t*-tests, to test for differences in disagreement - Fligner-Killeen Tests.

2.8.3 Survey details

Table 2.14: Selected survey questions from BOP-HH September 2020 wave

The inflation rate-Intro	
`Group filter: drandom1` $= 1 \| = 2$	
Now we would like you to think more carefully about the inflation rate.	
The inflation rate:	
Inflation is the percentage increase in the general price level. It is mostly measured using the consumer price index. A decrease in the price level is generally described as 'deflation'.	
Inflation development	
`Group filter: drandom1` $= 1 \| = 2$	`Group filter: drandom1` $= 3 \| = 4$
903A: What do you think the rate of inflation or deflation in Germany was over the past twelve months?	**903B:** By what percent do you think prices in Germany, in general, have increased or decreased over the past twelve months?
Note: If it is assumed that there was deflation, please enter a negative value.	Note: If it is assumed that prices have fallen, please enter a negative value.
Values may have a maximum of one decimal separator.	Values may have a maximum of one decimal separator.
Please use a full stop rather than a comma as the decimal separator.	Please use a full stop rather than a comma as the decimal separator.
Please enter a value here:	Please enter a value here:
`[Input field]` percent	`[Input field]` percent

Table 2.14: Selected survey questions from BOP-HH September 2020 wave (Continued)

`Group filter: drandom1` $= 1\| = 2$	`Group filter: drandom1` $= 3\| = 4$
005A1: Do you think inflation or deflation is more likely over the next twelve months?	**005A2:** Do you think prices in general, are more likely to increase or decrease over the next twelve months?
Note: Inflation is the percentage increase in the general price level. It is mostly measured using the consumer price index. A decrease in the price level is generally described as 'deflation'.	
Please select one answer:	Please select one answer:
1 = Inflation more likely	1 = More likely to increase
2 = Deflation more likely	2 = More likely to decrease
Inflation expectations quantitative	
`Group filter: drandom1` $= 1\| = 2$	`Group filter: drandom1` $= 3\| = 4$
`If 005A1` $= 1\| - 9997\| - 9998$	`If 005A2` $= 1\| - 9997\| - 9998$
005B1: What do you expect the rate of inflation in Germany to roughly be over the next twelve months?	**005B2:** By roughly what percentage do you expect prices in general, to increase over the next twelve months?
`If 005A1` $= 2$	`If 005A2` $= 2$
005B1: What do you expect the rate of deflation in Germany to roughly be over the next twelve months?	**005B2:** By roughly what percentage do you expect prices in general to decrease over the next twelve months?
Note: Inflation is the percentage increase in the general price level. It is mostly measured using the consumer price index. A decrease in the price level is generally described as 'deflation'.	

Table 2.14: Selected survey questions from BOP-HH September 2020 wave (Continued)

Please enter a value in the input field (values may have one decimal place). `[Input field]` percent	Please enter a value in the input field (values may have one decimal place). `[Input field]` percent
Inflation expectations qualitative	
`Group filter: drandom1` $= 1 \vert = 2$	`Group filter: drandom1` $= 3 \vert = 4$
005A1: Do you think inflation or deflation is more likely over the next twelve months? Note: Inflation is the percentage increase in the general price level. It is mostly measured using the consumer price index. A decrease in the price level is generally described as 'deflation'.	**005A2:** Do you think prices in general, are more likely to increase or decrease over the next twelve months?
Please select one answer: 1 = Inflation more likely 2 = Deflation more likely	Please select one answer: 1 = More likely to increase 2 = More likely to decrease
Inflation expectations quantitative	
`Group filter: drandom1` $= 1 \vert = 2$	`Group filter: drandom1` $= 3 \vert = 4$
`If 005A1` $= 1 \vert -9997 \vert -9998$	`If 005A2` $= 1 \vert -9997 \vert -9998$
005B1: What do you expect the rate of inflation in Germany to roughly be over the next twelve months?	**005B2:** By roughly what percentage do you expect prices in general, to increase over the next twelve months?
`If 005A1` $= 2$	`If 005A2` $= 2$
005B1: What do you expect the rate of deflation in Germany to roughly be over the next twelve months?	**005B2:** By roughly what percentage do you expect prices in general to decrease over the next twelve months?

Table 2.14: Selected survey questions from BOP-HH September 2020 wave (Continued)

Note: Inflation is the percentage increase in the general price level. It is mostly measured using the consumer price index. A decrease in the price level is generally described as 'deflation'.	
Please enter a value in the input field (values may have one decimal place).	Please enter a value in the input field (values may have one decimal place).
`[Input field]` percent	`[Input field]` percent
Inflation expectations probabilistic	
`Group filter: drandom1` = 1	`Group filter: drandom1` = 3
904A1: In your opinion, how likely is it that the rate of inflation will change as follows over the next twelve months?	**904A2:** In your opinion, how likely is it that prices in general will change as follows over the next twelve months?
Note: The aim of this question is to determine how likely you think it is that something specific will happen in the future. You can rate the likelihood on a scale from 0 to 100, with 0 meaning that an event is completely unlikely and 100 meaning that you are absolutely certain it will happen. Use values between the two extremes to moderate the strength of your opinion. Please note that your answers to the categories must add up to 100.	Note: The aim of this question is to determine how likely you think it is that something specific will happen in the future. You can rate the likelihood on a scale from 0 to 100, with 0 meaning that an event is completely unlikely and 100 meaning that you are absolutely certain it will happen. Use values between the two extremes to moderate the strength of your opinion. Please note that your answers to the categories must add up to 100.
The rate of deflation (opposite of inflation) will be 12% or higher	Prices will decrease by 12% or more
The rate of deflation (opposite of inflation) will be between 8% and 12%	Prices will decrease between 8% and 12%

Table 2.14: Selected survey questions from BOP-HH September 2020 wave (Continued)

The rate of deflation (opposite of inflation) will be between 4% and 8%	Prices will decrease between 4% and 8%
The rate of deflation (opposite of inflation) will be between 2% and 4%	Prices will decrease between 2% and 4%
The rate of deflation (opposite of inflation) will be between 0% and 2%	Prices will decrease between 0% and 2%
The rate of inflation will be between 0% and 2%	Prices will increase between 0% and 2%
The rate of inflation will be between 2% and 4%	Prices will increase between 2% and 4%
The rate of inflation will be between 4% and 8%	Prices will increase between 4% and 8%
The rate of inflation will be between 8% and 12%	Prices will increase between 8% and 12%
The rate of inflation will be 12% or higher	Prices will increase by 12% or more
`Group filter: drandom1` = 2	`Group filter: drandom1` = 4
904B1: What do you think the rate of inflation (or rate of deflation) is most likely to be over the next twelve months? What will the rate of inflation be as a maximum and minimum value?	**904B2** By what percentage do you think prices in general are most likely to increase or decrease over the next twelve months? What will the price change be as a maximum and minimum value?
Note: If it is assumed that there will be deflation, please enter a negative value.	Note: if it is assumed that prices will fall, please enter a negative value.
Values may have a maximum of one decimal place.	Values may have a maximum of one decimal place.
Please use a full stop rather than a comma as the decimal separator.	Please use a full stop rather than a comma as the decimal separator.

Table 2.14: Selected survey questions from BOP-HH September 2020 wave (Continued)

Most likely inflation or deflation rate `[Input field]` percent	Most likely change `[Input field]` percent
Minimum `[Input field]` percent	Minimum `[Input field]` percent
Maximum `[Input field]` percent	Maximum `[Input field]` percent

3 Quantifying Subjective Uncertainty in Survey Expectations[1]

3.1 Introduction

Expectations uncertainty matters in economics. Consumers who experience high inflation uncertainty, especially in times of economic turmoil, increase their savings (Armantier et al., 2021). Uncertain firms tend to respond less to monetary or fiscal policy (Bloom, 2009). Monitoring inflation expectations and the associated uncertainty may help recognize early signs of eroding central bank credibility or de-anchoring of inflation expectations (Grishchenko et al., 2019); central banks are paying increasing attention to consumer and firm expectations for this purpose (ECB, 2019). Subjective uncertainty also features prominently in theoretical models of expectation formation, such as rational inattention (Mackowiak and Wiederholt, 2009; Sims, 2003).

There is hence much interest in measuring uncertainty, both at the level of the aggregate economy (e.g. Baker et al., 2016; Carriero et al., 2018) and at the level of individual persons or firms. In the present paper, we propose a new measure of individual-level uncertainty based on reported subjective probabilities. Such a measure is an important input to studies that consider either the determinants or the consequences of subjective uncertainty. See, for example, Coibion et al. (2018a) for an analysis of firms' expectations, Ben-David et al. (2019) for a household finance perspective, and Clements et al. (2023) for an overview of macroeconomic expert forecasts.

Manski (2004, 2018) review a growing number of economic surveys in which participants assess the probability of a variable falling into various outcome ranges. In macroeconomics, the Survey of Professional Forecasters (SPF; Croushore, 1993) and its

[1] In press, corrected proof available online at `https://doi.org/10.1016/j.ijforecast.2023.06.001`

European counterpart (Garcia, 2003) are popular data sources covering expert forecasts. Furthermore, several surveys address the probabilistic expectations of consumers and firms. Examples include the Survey of Consumer Expectations (SCE) launched by the Federal Reserve Bank of New York (Armantier et al., 2017), a similar initiative by the Bank of Canada (Gosselin and Khan, 2015), and the firm survey by Coibion et al. (2018a). These data on probabilistic expectations promise to shed new light on consumers' uncertainty, complementing more traditional surveys using point expectations. The latter do not contain direct information about uncertainty. However, Binder (2017) utilizes a rounding pattern in the point forecasts data, namely respondents reporting multiples of five, in order to construct a measure of individual uncertainty.

Figure 3.1 illustrates subjective probability distributions ('histograms') from the April 2020 wave of the SCE.[2] Each survey participant provides probabilities for various outcome ranges ('bins') of next year's inflation rate, as represented by the horizontal axis. The SCE contains a substantial share of responses using only one or two bins. Such responses, called 'sparse histograms', are made by roughly a third of the SCE participants. Sparse histograms pose a challenge for existing measures of individual uncertainty (notably Engelberg et al., 2009, henceforth EMW) which are based on fitting a parametric distribution. For sparse histograms, fitting a flexible distribution is not possible, and a simple triangular shape is commonly used instead (see the two examples in the top row of Figure 3.1).

Motivated by the SCE data, we propose a new uncertainty measure that is transparent, trivial to implement, and well-defined even for sparse histograms. By contrast, existing approaches require assumptions on the support of the subjective histogram, the distribution within each bin, or the functional form of the underlying continuous distribution. Our proposed measure can be theoretically motivated as the generalized entropy function of the ranked probability score (Epstein, 1969), a strictly proper scoring rule. We, therefore, refer to the new measure as ERPS, for Expected Ranked Probability Score.

[2]Source: *Survey of Consumer Expectations*©, 2013-2023 Federal Reserve Bank of New York (FRBNY). The *SCE* data are available without charge at `http://www.newyorkfed/microeconomics/sce` and may be used subject to license terms posted there. FRBNY disclaims any responsibility for this analysis and interpretation of *Survey of Consumer Expectations* data.

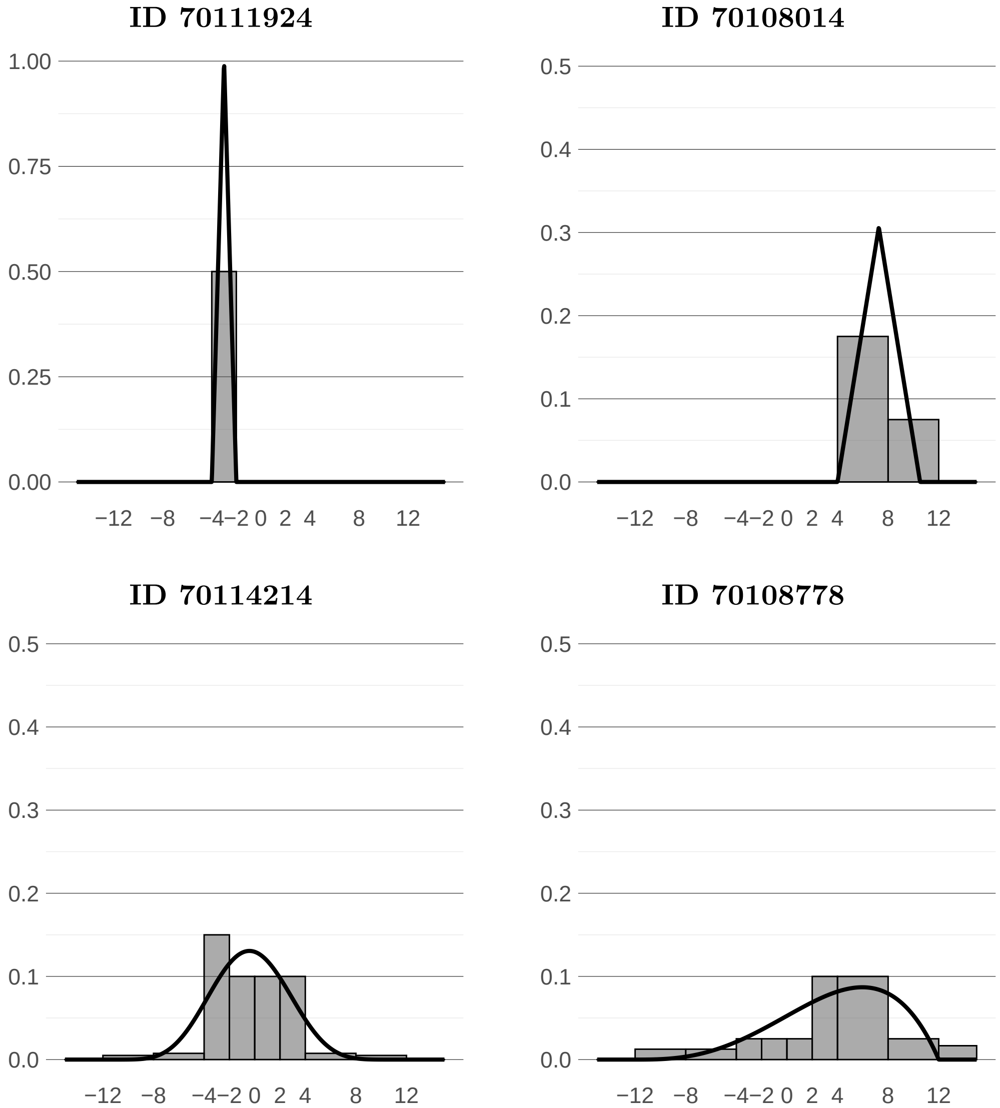


Figure 3.1: Illustration of probabilistic inflation expectations from the April 2020 wave of the SCE. The area of a rectangle corresponds to the subjective probability of the corresponding outcome range. For example, in the bottom left panel, the probability for an outcome between 2 and 4 equals $2 \times 0.1 = 0.2$. Solid lines indicate fitted probability density functions via the EMW method.

The remainder of this paper is structured as follows. Section 3.2 summarizes some stylized facts of the SCE probabilities. Section 3.3 describes existing methods for quantifying uncertainty. Section 3.4 develops the ERPS, detailing its advantages as mentioned above. Sections 3.5 and 3.6 study the behavior of the ERPS for simulated and empirical data, respectively. Section 3.7 concludes. The appendix contains details, proofs, and additional results.

3.2 Subjective Probabilities in the SCE Data

The SCE is conducted monthly with a sample size of about 1,300 respondents per month. The core module of the SCE asks, among others, for subjective probabilities of various outcome ranges, covering three variables: the inflation rate at two different horizons, real estate prices, and the respondent's personal earnings. In the SCE questionnaire made available by Federal Reserve Bank of New York (2020), the relevant question codes are Q9 and Q9c (inflation rate), C1 (growth rate of the average home price nationwide), and Q24 (growth rate of the respondent's personal earnings). The relevant outcome ranges (in percent), which are the same for all variables, can be represented by the intervals

$$(-\infty, -12]; (-12, -8]; (-8, -4]; (-4, -2]; (-2, 0]; (0, 2]; (2, 4]; (4, 8]; (8, 12]; (12, \infty).$$

These outcome ranges are depicted in the horizontal axis labels of Figure 3.1. In the case of inflation, for example, the two rightmost intervals refer to an inflation rate between 8% and 12% and to an inflation rate of 12% or more.[3]

Table 3.1 compares the SCE to expert forecasts in the SPF in terms of response behavior. The table's upper panel presents summary statistics on the number of histogram bins SCE participants used (the number of bins containing strictly positive probability mass). We focus on the time period from January 2014 to March 2020 for comparability to the SPF (see below). For inflation and the average home price, around 30% of the participants uses one or two bins ('sparse histograms'). For personal earnings, roughly

[3]The inclusion (or exclusion) of interval limits is not specified by the SCE survey questions. For example, the survey question leaves it unspecified whether an inflation rate of exactly 12% belongs to the last or penultimate bin. Our choice of half-open intervals is arbitrary – as is any choice in that regard – but seems unlikely to be of empirical relevance.

Table 3.1: Summary statistics on the number of bins used in the SCE (January 2014 to March 2020 waves) and SPF (2014:Q1 to 2020:Q1 waves); n denotes the total number of responses. We exclude histograms that do not sum to one (less than 0.4% of responses in both surveys).

		Share of respondents using			Mean nr.
	n	one bin	two bins	outer bin(s)	of bins
			SCE		
Average Home Price	85155	16.3	16.0	39.1	4.2
Inflation (one-year)	97019	12.6	17.3	38.9	4.4
Inflation (three-year)	97213	13.1	17.9	38.5	4.4
Personal Wage	65240	26.8	24.1	28.3	3.3
			SPF		
Inflation (GDP def.)	843	2.0	13.6	15.5	4.5
GDP	875	3.1	19.3	6.5	4.5
Inflation (CPI)	842	1.3	14.4	12.7	4.6
Inflation (PCE)	803	0.9	15.4	13.4	4.6
Unemployment	834	8.9	34.4	52.3	3.1

half of the participants use one or two bins. The mean number of bins used is higher for inflation and the average home price ($4.2 - 4.4$), than personal earnings (3.3). Finally, over a quarter of the participants use one or both of the outer bins that correspond to the intervals $(-\infty, -12]$ and $(12, \infty)$.

The lower panel of Table 3.1 presents analogous statistics for the SPF. The SPF questions are similar in design to those of the SCE, except that the two surveys use different numerical ranges for the histogram bins. While the SPF's bin definitions have been adapted over time (Federal Reserve Bank of Philadelphia, 2022), they are constant over the time period reported in Table 3.1. The number of bins (ten) is the same as in the SCE, except for GDP (eleven). While the share of participants using two bins and the mean number of bins used are comparable, there are some major differences to the SCE: First, the SPF features a much smaller share of participants who use a single bin. For example, this share is about ten percentage points lower for the inflation variables. Second, for the inflation variables, the share of participants using at least one outer bin is much lower in the SPF than in the SCE. Unemployment is the only SPF variable in

Table 3.1 for which participants make active use of an outer bin. However, this finding appears quite distinctive and is driven by a mismatch between the SPF's bin definitions and the empirical unemployment rate during the sample period considered in Table 3.1.[4]

Given its large sample size and the empirical patterns just reported, the SCE necessarily contains some histograms with non-standard shapes that are hard to capture by parametric distributions. Examples include distributions with multiple modes, distributions with 'holes' (strictly positive probability assigned to non-adjacent bins), or substantial probability mass in one or both outer bins. These features call for simple and robust methods that allow quantifying the uncertainty in any possible histogram.

3.3 Existing Uncertainty Measures

Survey probabilities as in Figure 3.1 do not specify a full probability distribution since the endpoints of the histogram's support as well as the distribution within each bin are unknown. Based on the raw probabilities alone, it is impossible to compute each participant's subjective mean or variance. In the following, we briefly review two methods that use parametric assumptions to account for the missing information.

3.3.1 Distribution Fitting

Following earlier work by Dominitz and Manski (1997), Engelberg et al. (2009, EMW) propose to fit a continuous distribution to the histogram probabilities. Their choice of continuous distribution depends on the number of histogram bins being used: EMW propose fitting a simple triangular distribution if the histogram is sparse and fitting a flexible generalized Beta distribution if the forecaster uses three or more bins. If the forecaster uses the leftmost bin (left limit of $-\infty$) or rightmost bin (right limit of $+\infty$), EMW propose treating the limits of the distribution's support as a free parameter. We provide details on the EMW method in Appendix 3.8.1. The method is used to derive

[4] As documented by Federal Reserve Bank of Philadelphia (2022), the bins for unemployment range from 'less than 4%' to 'more than 9%' for the sample period in question. Given that the actual US unemployment rate was close to or below 4% during much of the second half of the sample period, survey participants' use of the left outer bin seems empirically plausible. In retrospect, the SPF's bin definitions seem at odds with the empirical unemployment rate. Indeed, the SPF bin definitions were changed from 2020:Q2 onwards, reflecting a wider range of unemployment outcomes.

uncertainty measures that are reported in official SCE publications such as Armantier et al. (2017), and are made available for download by Federal Reserve Bank of New York (2020).

The EMW method provides a full analytical distribution from which any feature of interest (such as subjective measures of location, spread, or tail risk) can be computed. However, this wealth of information comes at a cost: First, choosing a particular parametric distribution seems hard to justify for sparse histograms and is potentially restrictive even for dense histograms. For example, the generalized Beta distribution cannot accommodate multimodal histograms, which may be empirically relevant in some situations. (In principle, the generalized Beta distribution could accommodate two modes at the left and right end of the support. However, this type of bimodality seems empirically implausible, and Engelberg et al. propose to exclude it when fitting the distribution. See Appendix 3.8.1 for details.) Second, the approach entails a discontinuity when moving from a histogram with two bins (approximated via a triangular distribution) to one with three bins (approximated via a generalized Beta distribution). Finally, practical implementation requires judgmental choices pertaining, e.g., to parameter limits imposed in numerical optimization, or to the handling of certain 'undefined' cases that are not covered by EMW's proposal (because they did not or rarely occur in their SPF data) but that inevitably occur in large data sets like the SCE. Such implementation choices may reasonably be made differently by different authors. Full reproducibility hence requires careful documentation of all choices.

For the SPF data, the drawbacks of the EMW method arguably play a minor role since both the share of sparse histograms and the share of 'undefined' cases are small. This observation explains the widespread and successful use of the EMW method for the SPF and similar data sets. By contrast, given the properties of the SCE discussed above, the EMW method seems less well-adapted to large-scale consumer surveys.

3.3.2 Mass-at-midpoint method

The mass-at-midpoint (MAM) method (see Glas, 2020, and the references therein) assumes that the subjective distribution is discrete, with a point mass at $\{m_k\}_{k:p_k>0}$, where m_k denotes the midpoint of bin $k = 1, \ldots, K$. Hence the method assumes point mass at the subset of bins that receive nonzero probability. Under this assumption, the subjective mean and standard deviation can easily be computed. An advantage of this method is that it can be applied irrespective of the number of bins used. In particular, it avoids the discontinuity inherent in the EMW method. A disadvantage of the MAM method arises whenever the participants use one of the outer bins (i.e., whenever $p_1 > 0$ or $p_{10} > 0$). In this case, the subjective mean and standard deviation depend on the endpoints of the outer bins, which are not specified by the survey design and for which assumptions seem hard to justify. This disadvantage is especially relevant for the SCE, where about one-third of the participants use at least one outer bin. We provide evidence on this aspect in Appendix 3.8.4.

3.4 A New Approach to Quantifying Uncertainty in Survey Histograms

3.4.1 General Idea: Quantifying Uncertainty via Entropy

We treat each survey response as a vector of probabilities $\underline{p} := \left(p_1, \quad p_2, \quad \ldots, \quad p_K\right)'$, where p_k denotes the subjective probability that the inflation rate is within the interval r_k that defines the range of bin k. In practice, the intervals $\{r_k\}_{k=1}^{K}$ are disjoint and their union is the real line. Hence the probabilities $\underline{p}$ form a subjective survey histogram as in Figure 3.1.

Our proposed measure of uncertainty is based on the concept of entropy. Informally, if the entropy of a distribution $\underline{p}$ is large, then a forecaster with subjective distribution $\underline{p}$ places a high probability on making large forecast errors. In that sense, $\underline{p}$ corresponds to high uncertainty. Vice versa, under a low-entropy distribution $\underline{p}$, large forecast errors are unlikely, and hence low entropy corresponds to low uncertainty.

More formally, entropy relates to strictly proper scoring rules (Gneiting and Raftery, 2007). In economics, scoring rules are commonly used for eliciting beliefs in experiments (Schotter and Trevino, 2014) and for evaluating probabilistic forecasts (e.g. Boero et al., 2011). In a discrete setup, scoring rules are functions of the form $\mathrm{S}(\underline{p}, k^*)$ that measure the performance of the probabilistic forecast $\underline{p}$ if the outcome k^* realizes. The integer $k^* \in \{1, 2, \ldots, K\}$ indicates the histogram bin that contains the realization. We consider specific choices of S below. For each of these choices, a smaller value of S indicates a better forecast. A scoring rule S is called strictly proper if a forecaster minimizes their expected score by stating what they think is the true probability distribution $\underline{p}$ (conditional on their information set); see Gneiting and Katzfuss (2014, Section 3.1.1) for a formal definition. The function

$$\mathrm{ES}(\underline{p}) \quad = \quad \sum_{k=1}^{K} p_k \; \mathrm{S}(\underline{p}, k)$$

is called the entropy function associated with the scoring rule S (e.g. Gneiting and Raftery, 2007, Section 2.2). We propose to use this function to measure the subjective uncertainty in a probabilistic survey forecast $\underline{p}$.

3.4.2 Expected Ranked Probability Score (ERPS)

As our preferred choice of scoring rule S, we consider the ranked probability score (RPS; Epstein, 1969):

$$\mathrm{RPS}(\underline{p}, k^*) \quad = \quad \begin{cases} \sum_{k=1}^{K}(1-P_k)^2 & \text{if } k^* = 1 \\ \sum_{k=1}^{k^*-1}(P_k)^2 + \sum_{k=k^*}^{K}(1-P_k)^2 & \text{if } k^* \in \{2, 3, \ldots, K\}, \end{cases}$$

where $P_k = \sum_{j=1}^{k} p_j$ is the cumulative probability of the first k bins. As its name suggests, the RPS is designed for ranked categorical variables. That is, the RPS treats the realizing bin $k^* \in \{1, \ldots, K\}$ as an ordinal variable, with $k^* = 1$ representing a smaller outcome of the underlying variable than $k^* = 2$.[5] Thus, the RPS rewards forecasters who put much probability mass into bins equal to or close to the realizing bin k^*. For example, if a forecaster places unit probability mass on the third bin, then $k^* = 2$ yields a lower (i.e., better) RPS than $k^* = 1$. Boero et al. (2011) persuasively argue that this feature of the RPS is well in line with survey histograms, and propose to use it for evaluating the histograms' predictive accuracy.

[5] In our empirical analysis based on the SCE's bin definitions, the first bin, $k = 1$, ranges from $(-\infty, -12]$, the second bin, $k = 2$, ranges from $(-12, -8]$, and similarly for the other bins. The last bin, $k = 10$, represents outcomes in the range $[12, \infty)$.

The entropy function of the RPS is given by

$$\begin{aligned}\mathrm{ERPS}(\underline{p}) &= \sum_{k=1}^{K} p_k \ \mathrm{RPS}(\underline{p}, k) \\ &= \sum_{k=1}^{K} P_k(1 - P_k). \qquad (1)\end{aligned}$$

The latter equation, our proposed uncertainty measure, is trivial to compute from the histogram probabilities.

Since it attaches only an ordinal but not a numerical interpretation to the bins, the ERPS at (1) does not depend on the bins' outcome ranges or the (unknown) distribution of probability mass within each bin. The ordinal interpretation renders parametric assumptions obsolete and explains the simplicity and robustness of the ERPS. For example, the ERPS easily accommodates sparse or multimodal histograms. A drawback of the ordinal interpretation is that the ERPS is not comparable across different bin definitions, such as design A involving ten bins of length one and design B involving five bins of length two covering the same interval. This concern may be relevant if the bin definitions must be adapted over time to account for changes in the distribution of the predictand. Such re-definitions occurred several times for the SPF since it was launched in 1968 (see Federal Reserve Bank of Philadelphia, 2022). However, the concern is less relevant for the SCE whose probability ranges have remained unchanged since its start in 2013, and have also been adopted by a number of other consumer surveys such as the Bundesbank's Online Panel (Deutsche Bundesbank, 2022).

3.4.3 Comparison to Other Entropy-based Measures

Here we relate the ERPS to entropy functions for two other popular scoring rules. The logarithmic score (LS; Good, 1952) and Brier score (BS; Brier, 1950) are given by

$$\begin{aligned}\mathrm{LS}(\underline{p}, k^*) &= -\log p_{k^*} \\ \mathrm{BS}(\underline{p}, k^*) &= \sum_{k=1}^{K} (\mathbb{I}_{k=k^*} - p_k)^2,\end{aligned}$$

where $\mathbb{I}_{k=k^*}$ is an indicator function that equals one if $k = k^*$, and equals zero otherwise. Their respective entropy functions are given by

$$\begin{aligned}
\text{ELS}(\underline{p}) &= -\sum_{k=1}^{K} p_k \log p_k. \\
\text{EBS}(\underline{p}) &= \sum_{k=1}^{K} p_k(1 - p_k).
\end{aligned}$$

The ELS was famously developed by Shannon (1948) and is typically called 'Shannon Entropy'. In economics, it plays a key role in the theory of rational inattention (Sims, 2003). Rich and Tracy (2010) use the ELS to measure uncertainty of the SPF histograms. The EBS is much less widely used, with the interesting exception of López-Menéndez and Pérez-Suárez (2019) who quantify uncertainty in (aggregate) tendency surveys.

The BS and LS are designed for multinomial random variables; the outcome categories $k^* \in \{1, \ldots, K\}$ are considered interchangeable. Hence the EBS and ELS are invariant to permutations of the histogram probabilities $p_1, \ldots, p_K$. For example, for a hypothetical three-bin histogram, the probabilities $\underline{p}_a = (1/4, 1/2, 1/4)'$ yield the same EBS as the probabilities $\underline{p}_b = (1/2, 1/4, 1/4)'$. This assessment seems implausible, given that $\underline{p}_b$ is obtained from $\underline{p}_a$ by shifting probability mass from the central bin to the more extreme leftmost bin. Under the ERPS, which utilizes an ordinal interpretation, $\underline{p}_b$ is considered more uncertain than $\underline{p}_a$.

ELS and EBS are both maximized by the vector

$$\underline{p}^{**} = \tau \times (1/K),$$

where τ is a $K \times 1$ vector of ones (see López-Menéndez and Pérez-Suárez 2019, Shannon 1948). Hence flat probabilities represent maximal uncertainty, as seems natural in a multinomial setup. By contrast, we show in Appendix 3.8.2 that the maximal ERPS is attained for the vector

$$\underline{p}^* = \Big(1/2, \quad 0, \quad \ldots, \quad 0, \quad 1/2\Big)'$$

that places probability one-half on each of the two outer bins. The intuition for this solution is that under $\underline{p}^*$, it is certain that one of the two outer bins will materialize.

Both outcomes produce a large score RPS($\underline{p}^*, k$), since $\underline{p}^*$ places no probability mass on the neighboring bins.

While the RPS accounts for ordering of the outcome categories $k^* \in \{1, \ldots, K\}$, it does not reflect information about the width of the corresponding histogram bins. This information requires a numerical, rather than just ordinal, interpretation of the outcome categories. As mentioned, the numerical interpretation is challenging in the present context. Nevertheless, relating the (E)RPS to entropy-based uncertainty measures for numerical outcomes is interesting. We provide such a comparison in Appendix 3.8.3.

3.5 Simulation Studies

This section compares the ERPS to the EMW and MAM methods of quantifying survey uncertainty.

3.5.1 Survey Histograms as Noisy Realizations

Our first simulation design views survey histograms as a noisy realization of an underlying true continuous distribution. Survey noise could arise, for example, from participants' limited attention when answering the survey. In the following, we analyze which histogram-based uncertainty measure is most closely aligned with the uncertainty of the true distribution. We implement this idea via the following design:

- Draw an independent sample of size n from a random variable X with continuous distribution F
- Set the 'survey' probability for the jth bin equal to

$$\hat{p}_j = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(x_i \in \text{bin}_j),$$

  where $\mathbf{1}(A)$ is the indicator function of the event A, x_i is the ith realization of the simulated sample, and bin_j is the j-th SCE interval as defined in Section 3.2, with $j \in \{1, 2, \ldots, 10\}$.

- Denote the corresponding true probability for bin j by

$$p_j = \mathbb{P}(X \in \text{bin}_j) = \int_{\text{bin}_j} \mathrm{d}F(x).$$

We next compute various uncertainty measures based on the simulated survey probabilities $\{\hat{p}_j\}_{j=1}^{10}$ and, possibly, the corresponding bin limits. We then compare these measures to the underlying ground truth measure of uncertainty. Specifically, for the EMW-SD and MAM-SD methods, we compare the estimated standard deviation $\hat{\sigma}$ to σ, the true standard deviation implied by F.[6] Similarly, we compare the interquartile range estimated by the EMW method (EMW-IQR) to the true interquartile range implied by F, and we compare the estimated ERPS (based on the $\hat{p}_j$'s) to the true ERPS (based on the p_j's).

The degree of noise in the histograms is governed by n. While a small sample size n may entail large deviations between $\hat{p}_j$ and p_j (and, possibly, empty bins $\hat{p}_j = 0$ for some j), each $\hat{p}_j$ converges in probability to p_j as $n \to \infty$. It remains to choose a true distribution F for simulating the data. We consider two variants: First, a Gaussian distribution $\mathcal{N}(\mu, \sigma)$, where, μ and σ are the quantified means and standard deviation associated with a randomly selected SCE histogram, drawn from June 2013 to April 2020 waves and requiring that the histogram uses at least two bins. Second, the EMW method's quantified (triangular or generalized Beta) distribution for a randomly selected SCE histogram, again requiring at least two bins. While the first variant is somewhat simpler, the second variant intentionally favors the EMW method, because it simulates data that by construction are closely in line with its functional form assumptions.

Table 3.2 summarizes the results, across 1 000 Monte Carlo iterations and for two sample sizes $n \in \{20, 50\}$. Reassuringly, there is generally a high agreement between the estimated and the true uncertainty values, with all rank correlations in Table 3.2 exceeding 0.94. As expected, all methods perform better in the setup with less noise ($n = 50$). While the methods perform very similarly in the Gaussian case, the second variant indicates a modest advantage of the EMW and ERPS methods as compared to MAM-SD.

[6] In both simulation studies, we implement the MAM method by assuming that the left and right end of the histogram's support is given by -16 and 16, respectively. We provide further evidence on this implementation choice in Section 3.6 below.

Table 3.2: Spearman rank correlation between estimated and true uncertainty across 1 000 Monte Carlo simulations.

	Normal		Quantified (triangular or gen. Beta)	
	$n = 20$	$n = 50$	$n = 20$	$n = 50$
EMW-SD	0.97	0.99	0.97	0.99
MAM-SD	0.95	0.97	0.95	0.96
EMW-IQR	0.97	0.98	0.97	0.99
ERPS	0.97	0.98	0.97	0.99

Given that the second variant favors EMW uncertainty measures by construction, the good performance of the ERPS confirms its robustness.

3.5.2 Sparse Histograms

As we have argued, a key advantage of the ERPS over the EMW method is that the former requires no case distinction when moving from a sparse histogram (using two bins) to a histogram using three bins. We demonstrate the quantitative relevance of this point in a simulation study based on June 2013 to April 2020 waves of the SCE. For comparability, we focus on participants who use two adjacent bins, none of which is an outer bin in the SCE's histogram design shown in Section 3.2. We further require that the histogram probabilities sum to one and exceed one percent, which is the magnitude of the perturbation we consider. These selection criteria leave us with 15 961 two-bin histograms. For each histogram, we consider two simple perturbations: First, we move one percentage point of probability mass from the left bin to its left neighboring bin. For example, suppose that the original histogram allocates 50% probability to each of the two bins $(0, 2]$ and $(2, 4]$. The perturbed histogram then places probability $1\%, 49\%$ and 50% to the three bins $(-2, 0]$, $(0, 2]$ and $(2, 4]$ respectively. Second, we apply an analogous perturbation to the right histogram bin, such that the perturbed histogram contains one percent of probability mass in a third bin located to the right of the original histogram. We choose a perturbation size of one percentage point since it is the smallest size that seems empirically plausible.

For each setup (no perturbation, left perturbation, and right perturbation), we again consider the ERPS, as well as the standard deviation (EMW-SD) and interquartile range (EMW-IQR) of the distribution obtained via the EMW method, and the standard deviation obtained via the MAM method (MAM-SD). Given the small perturbation size, we contend that an uncertainty metric should be robust to the perturbation.[7] To measure the similarity between the perturbed and baseline histograms, we consider the rank correlation between the uncertainty measures and their mean absolute deviation (MAD). Table 3.3 summarizes the results, indicating that the perturbation significantly impacts the two EMW measures. The rank correlation between the original and perturbed measures can be as low as 0.38, which is remarkable given the small magnitude of the change. Similarly, the mean absolute deviations in the first two rows of Table 3.3 are considerable given the mean values of the uncertainty measures reported in the first column. The results further indicate that the impact of right perturbation is larger than the impact of the left perturbation. This effect is due to the empirical pattern that many of the two-bin histograms focus on the bins $(2, 4]$ and $(4, 8]$. According to the SCE's bin design shown in Section 3.2, the left neighbor of these bins is at $(0, 2]$, whereas the right neighbor is at $(8, 12]$. Hence, the left perturbation expands the support of the histogram by two units, whereas the right perturbation expands the support by four units. This asymmetry matters here since the Engelberg et al. (2009) algorithm supports the histogram if only interior bins are used.

MAM-SD is robust to both left and right perturbation, attaining rank correlations close to one and reducing mean absolute deviations by about 50% compared to the EMW method. For ERPS, the impact of the perturbation can be described analytically. Let $\underline{p}$ denote a two-bin histogram, and $\tilde{\underline{p}}_L$ and $\tilde{\underline{p}}_R$ its perturbed version with probability mass shifted to the left and right neighboring bin, respectively. Let δ denote the perturbation size (with $\delta = 0.01$ in our simulation study). Then Eq. (1) yields that

$$\mathrm{ERPS}(\tilde{\underline{p}}_L) = \mathrm{ERPS}(\tilde{\underline{p}}_R) = \mathrm{ERPS}(\underline{p}) + \delta\ (1 - \delta),$$

[7]For larger perturbation sizes, it is no longer clear whether the uncertainty measure should be robust to the perturbation or not.

Table 3.3: First column: Mean of uncertainty measure (without perturbation). Second to fifth column: Rank correlation of uncertainty in perturbed and baseline histograms, and mean absolute deviation (MAD) between uncertainty in perturbed and baseline histograms.

		Left perturbation		Right perturbation	
	Mean	Correlation	MAD	Correlation	MAD
EMW-SD	0.75	0.66	0.10	0.38	0.19
EMW-IQR	1.08	0.70	0.13	0.54	0.24
MAM-SD	1.00	0.99	0.05	0.97	0.09
ERPS	0.19	1.00	0.01	1.00	0.01

i.e. both perturbations lead to an additive increase in ERPS by $\delta\ (1-\delta)$. Hence, the perturbation affects all histograms in exactly the same way, leading to correlations of one and mean absolute deviations of about 0.01, demonstrating the robustness of the ERPS.

The present simulation experiment aims to quantify the impact of the EMW method's known discontinuity when moving from two to three bins. The two-bin case is empirically relevant for individual-level SCE histograms, but less common in other contexts (such as average histograms across many survey participants). When the baseline histogram uses three or more bins, we expect the EMW method to be reasonably robust to the types of small perturbations considered above, as the switch from a triangular to a generalized Beta distribution for quantification does not occur in these situations.

3.6 Empirical Comparisons

We next compare the ERPS to the EMW and MAM methods based on empirical survey data. For EMW, we focus on the EMW-SD variant; the results based on EMW-IQR are qualitatively identical and are hence omitted for brevity. For MAM, we initially assume that the lower and upper support limits are given by -16 and 16, such that the open bins have the same length as their neighboring closed bins. We then study the impact of this parameter at the end of the section and in Appendix 3.8.4.

Overall, the three uncertainty measures display strong positive associations. For a given survey date and variable, the rank correlation between any two uncertainty measures

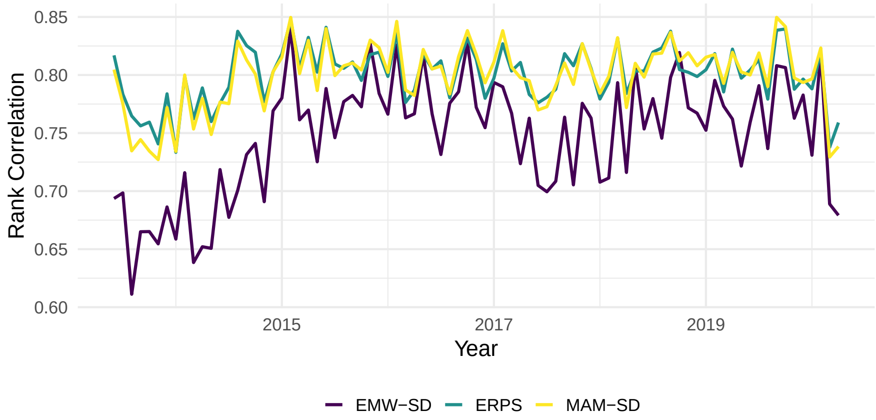


Figure 3.2: Rank correlation of subjective uncertainty regarding house prices and inflation (one year ahead). The lines correspond to different measures of uncertainty.

is at least 0.87, and often as high as 0.95. We next analyze whether respondents who express high uncertainty about inflation also express high uncertainty about house prices and their personal earnings. To this end, we consider the rank correlation coefficient of uncertainty across variables. We consider six pairs of variables and 83 monthly survey waves (June 2013 to April 2020). Figure 3.2 illustrates house prices and inflation, indicating that ERPS and MAM generally yield a higher rank correlation than EMW-SD. Similar patterns also hold more broadly: Across all variable pairs and survey dates, the ERPS attains the highest rank correlation in about 64% of all cases. The corresponding shares for MAM and EMW are 34% and 2%. There is hence clear evidence that the ERPS is more consistent across variables than the EMW method. In the absence of a 'ground truth' measure of uncertainty, we cannot tell whether this feature of the ERPS is desirable. However, these findings indicate an interesting and robust difference between both measures.

We further compare the persistence of uncertainty as measured by EMW, ERPS, and MAM. We measure persistence by the rank correlation of uncertainty in two subsequent SCE waves, for the subset of participants present in both waves. A small rank correlation may indicate a genuine shift in relative uncertainty from one month to the next (e.g., Anne is more uncertain than Bob in January, whereas Bob is more uncertain than Anne

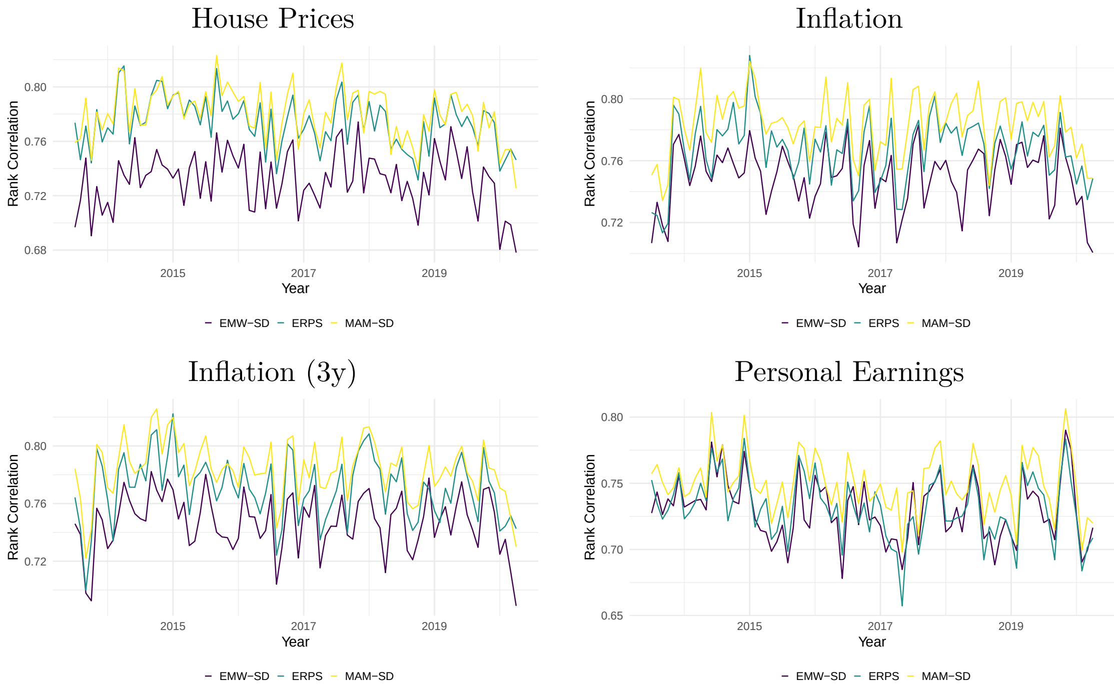


Figure 3.3: Rank correlation of subjective uncertainty over two subsequent survey months, based on participants present in both months.

in February). Alternatively, the small rank correlation may reflect noise in the uncertainty measure.

Figure 3.3 presents results on the persistence of uncertainty. The observed correlation of all three uncertainty measures is similar for personal earnings. Genuine shifts in relative uncertainty seem particularly plausible for this variable since it is individual-specific and prone to idiosyncratic information updates (such as Anne signing a new labor contract in February). MAM-SD is most persistent overall for house prices and inflation, followed by ERPS and EMW-SD. Similar to the findings across variables, this indicates that MAM-SD and ERPS are less sensitive to small changes in the raw probabilities $\underline{p}$ than EMW-SD.

Until now, all (simulation and empirical) results for MAM have assumed lower and upper support limits of -16 and 16. This choice is necessarily judgmental as the histograms yield no information on support limits. One could argue that support limits of ± 16 are too narrow, and consider wider limits instead. However, any particular choice of wider limits is arbitrary as well. In Appendix 3.8.4, we provide empirical results on this topic for inflation expectations in the SCE. As shown there, the largest standard deviations in the sample are especially sensitive to the choice of support limits and become very large for wide choices (such as ± 38 used in the 'wide' scenario of Appendix 3.8.4). On the other hand, these wide choices are not easily refuted as implausible, e.g. when considering the presence of extreme point expectations in consumer surveys.[8]

3.7 Discussion

This paper introduces the ERPS, a new measure of uncertainty in probabilistic survey expectations. The ERPS is based on an ordinal interpretation of the survey outcome categories which prevents parametric assumptions and explains its simplicity and robustness. The Engelberg et al. (2009, EMW) method, the current standard for quantifying uncertainty in economic surveys, uses a numerical interpretation of outcome categories instead. The numerical interpretation is more demanding and requires the researcher to make parametric assumptions about unknown aspects of the histogram. In return, it provides a full picture of subjective uncertainty.

We think that a user's choice between the ERPS and the EMW method should depend on the signal-to-noise ratio in the subjective probability data. If this ratio is high, then the EMW method – which is more sensitive to small changes in the probabilities – seems more appropriate. Examples of this situation include average histograms across time or across socio-demographic groups (which may be based on hundreds of individual responses), and perhaps probability assessments by individual expert forecasters. By contrast, the ERPS seems preferable in the context of individual-level probabilities by consumers, such as the ones covered by the SCE. This data type is an innovative source for monitoring and studying the general public's inflation expectations. In particular, microdata from the SCE and similar surveys allow us to analyze the heterogeneity in economic expectations across socio-demographic subgroups of society. Such analyses are

[8] Consider, for example, one year ahead inflation expectations in the SCE (variable code Q8v2part2). Across the June 2013 – April 2020 survey waves, 1.3% of the point expectations are -25% or lower, and 5% of the point expectations are $+25\%$ or higher.

highly relevant to study the general public's response to economic policy measures (see e.g. D'Acunto et al., 2023). Finally, for aggregate measures of uncertainty (obtained, e.g., by computing an individual-level measure of uncertainty, and then averaging this measure across survey participants), we typically expect the difference between the EWM method and the ERPS to be limited. Due to averaging, the sensitivity of the EMW method will often be of minor importance in such situations.

Our simulation and empirical results also cover the mass-at-midpoint (MAM) method, which can estimate the standard deviation corresponding to an individual-level histogram. While MAM seems more attractive than the EMW method for two-bin histograms (see Section 3.5.2), its use of judgmental support limits makes it less attractive for histograms involving outer bins (see Appendix 3.8.4). Compared to the ERPS, MAM enables more detailed information on a forecast histogram (in particular, estimates of its mean and standard deviation), at the cost of making assumptions and implementation choices that are hard to justify rigorously. This trade-off is similar to the trade-off that arises when comparing the ERPS and the EMW method.

Finally, while we have focused on measuring subjective uncertainty by itself, an interesting question is whether subjective uncertainty lines up with measures of realized uncertainty based on expectation errors. This comparison is of economic relevance since over– or underestimating objective uncertainty has possibly severe implications for decision making (see e.g. Ben-David et al., 2013). In Appendix 3.8.5, we demonstrate that the ERPS can also be used in this context.

3.8 Appendix

The appendix provides details on the EMW method (Appendix 3.8.1), proves a claim on the ERPS (Appendix 3.8.2), relates the ERPS to the CRPS (Appendix 3.8.3), investigates the choice of bin limits for the MAM method (Appendix 3.8.4), and sketches a comparison of the ERPS to its realized counterpart (Appendix 3.8.5).

3.8.1 Details on the EMW Method

Here we provide details on our implementation of the (Engelberg et al., 2009, EMW) method for quantifying forecast histograms.

Case A: Forecaster uses one or two bins

Following Engelberg et al. (2009, EMW), we construct isosceles triangles that are completely characterized by their support which we denote by $[a, b]$. The mode of the distribution is located at $c = (a + b)/2$.

If a forecaster uses only one bin, we use a triangular distribution with support equal to that of the bin used. This approach, which is recommended in EMW's Section 4.1.1, differs from the one implemented in the SCE, which assumes a uniform distribution over the support of the bin (Armantier et al., 2017, Footnote 28).

In the case of a forecaster using two adjacent bins, Becker et al. (2022) note that the original procedure by EMW may yield counterintuitive triangular fits when applied to survey probability intervals of varying widths (like the SCE). Suppose the two bins with nonzero probability are given by $[L, M)$ and $[M, R)$, and denote the corresponding probabilities by p_L and p_R. Similar to Becker et al., we set $a = L$, if the left interval features weakly higher density, i.e., if $p_L/(M - L) \geq p_R/(R - M)$. Otherwise, we set $b = R$. Unlike Becker et al. we then choose the other endpoint of the isosceles triangle by numerically optimizing the squared difference between the empirical and fitted CDFs. This approach is motivated by the fitting criterion used in the case of three or more bins, described below. In most empirical two-bin cases, a squared difference of zero is attainable, and the numerical solutions coincide with the formulas proposed by Becker et al.. However, exceptions to this situation exist, including the example of a participant placing 30% and 70% probability on the (2,4] and (4,8] bins, respectively. The preceding description does not cover two scenarios:

- The forecaster uses two non-adjacent bins such as $(0, 2]$ and $(4, 8]$.
- The forecaster uses one or two bins, including one of the outer bins (i.e., $p_1 > 0$ or $p_K > 0$).

The EMW method does not prescribe a solution for the former scenario. In the latter scenario, any solution would seem to hinge on an arbitrary choice of support limit. In our analysis, we drop observations from either of the two scenarios to not distort our findings on the EMW method.

Case B: Forecaster uses three or more bins

If the forecaster uses three or more bins, EMW propose to fit a generalized Beta distribution given by

$$\begin{aligned} F_{\text{gBeta}}(x; a, b, l, r) &= \begin{cases} 0 & x \leq l, \\ \frac{1}{B(a,b)} \int_l^x \frac{(u-l)^{a-1}(r-u)^{b-1}}{(r-l)^{a+b-1}} du & l < x \leq r, \\ 1 & x > r, \end{cases} \qquad (2) \\ B(a, b) &= \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \\ \Gamma(a) &= \int_0^\infty u^{a-1} \exp(-u)\ du. \end{aligned}$$

Instead of the limits 0 and 1 of the regular Beta distribution, F_{gBeta} entails flexible left and right limits $l, r \in \mathbb{R}$ with $l < r$. The two shape parameters $a, b \in \mathbb{R}_+$ play the same role as in regular Beta distributions. EMW impose the constraint that $a > 1$ and $b > 1$ to obtain a unimodal shape, which seems plausible in the present context.

To fit the distribution at (2) to a vector of histogram probabilities $\underline{p}$, EMW propose to fix the limits l and r at the endpoints of the bins that are being used. If one or both of the two outer bins are being used, the authors propose to treat the limits l and/or r as free parameters to be estimated. That is, l is a free parameter if $p_1 > 0$, and r is a free parameter if $p_K > 0$, where $K = 10$ in the case of the SCE. Following Armantier et al. (2017, Appendix C), we impose the constraint that $l > -38$ and that $r < 38$ when estimating l and/or r. We further impose that $l < -12$ and $r > 12$, as is logically required by the SCE's bin design. The shape parameters a and b are estimated in either case. In the most general case where l and r are both estimated, the fitting problem is thus given by

$$\max_{\substack{a>1, b>1,\\ -38<l<-12,\\ 12<r<38}} \sum_{k=1}^{K} \left[F_{\text{gBeta}}(x_k; a, b, l, r) - P_k\right]^2,$$

where x_k is the right endpoint of the kth histogram bin, and $P_k = \sum_{j=1}^{k} p_j$ is the cumulative probability of the first k bins.

R code for implementing the described quantification method just described is available via the first author's website. We drop a small number of individual survey responses (12, out of 367 728) for which our quantification method leads to excessive numerical challenges, primarily due to a large probability mass in one of the outer histogram bins.

3.8.2 Maximal ERPS

Here we prove a claim made in Section 3.4.3 of the paper.

The ERPS of a distribution $\underline{p}$ is given by

$$\text{ERPS}(\underline{p}) = \sum_{k=1}^{K} P_k(1 - P_k)$$

In matrix notation, let $\underline{p}$ be the $K \times 1$ vector with probabilities p_k, and $\underline{P}$ be the corresponding vector of cumulative probabilities P_k. We have that $\underline{P} = C'\underline{p}$, where C is a $K \times K$ upper triangular matrix with all elements above the main diagonal equal to one, and all diagonal elements equal to one. We can write

$$\text{ERPS}(\underline{p}) = \underline{P}'(\tau - \underline{P}) = \underline{p}'C\tau - \underline{p}'CC'\underline{p},$$

where τ is a $K \times 1$ vector of ones. To find the maximand of the ERPS, we solve the following problem:

$$\arg\max_{\underline{p}} \text{ ERPS}(\underline{p}) \text{ such that } \underline{p}'\tau = 1;$$

note that the constraint that probabilities be non-negative need not be enforced explicitly. Setting up the Lagrangian and solving the resulting quadratic problem then shows that the maximand is given by

$$\underline{p}^* = \Big(1/2, \quad 0, \quad \ldots, \quad 0, \quad 1/2\Big)';$$

note that the second-order condition for a maximum is satisfied since CC' is strictly positive definite.

3.8.3 Relating the ERPS to Probability Distributions for Numerical Outcomes

Summary

The ERPS attaches an ordinal interpretation to the histogram bins and depends only on the vector $\underline{p}$ of bin probabilities only. Here we relate this perspective to the unknown probability distributions for numerical outcomes that may underlie a given vector $\underline{p}$. In particular, consider two different probability distributions F_1, F_2, such that both F_1 and F_2 match $\underline{p}$, i.e., they both assign probability p_k to the interval defining bin $k = 1, \ldots, K$. F_1 and F_2 attain the same ERPS. By contrast, uncertainty measures for numerical outcomes will typically assign different levels of uncertainty to F_1 and F_2. In that sense, the ERPS summarizes the uncertainty of all probability distributions F that match $\underline{p}$. Below we discuss this conceptual aspect in more detail, providing explicit results for one particular subclass of distributions that match $\underline{p}$, and for the simplified case of a histogram with bins of equal width. The requirement of equal bin widths is natural and essential: Since the ERPS does not use information on bin length, there is no meaningful way to study the ERPS in a setup where bin length is a relevant parameter.

Details

Consider a discrete random variable X with support $v_1, v_2, \ldots, v_K$, where $v_j \in \mathbb{R}$ for all $j = 1, \ldots, K$, $v_a < v_b$ for $a < b$, $\mathbb{P}(X = v_j) = p_j$, and $\sum_{j=1}^K p_j = 1$. We think of X (and its modified versions below) as a draw from the probability distribution that underlies a given survey histogram. The cumulative distribution function (CDF) of X is given by

$$F(x) = \mathbb{P}(X \leq x) = \sum_{j:\ v_j \leq x} p_j.$$

Hence $F(x)$ is a piecewise constant function that satisfies $F(x) = 0$ for $x < v_1$ and $F(x) = 1$ for $x \geq v_K$. Since X is supported on the real line, we can measure its underlying uncertainty using the continuous ranked probability score (CRPS; Matheson and Winkler, 1976), a strictly proper scoring rule. Observe that using the CRPS requires a numerical interpretation of the support of X, in contrast to the ordinal interpretation underlying the ERPS. The expected CRPS (ECRPS), or CRPS entropy, for X is given by

$$\begin{aligned}
\text{ECRPS}(F) &= \int_{-\infty}^{\infty} \text{CRPS}(F, x)\ dF(x) &(3)\\
&= \int_{-\infty}^{\infty} F(x)(1 - F(x))\ dx \\
&= \int_{v_1}^{v_K} F(x)(1 - F(x))\ dx \\
&= \sum_{j=1}^{K-1} (v_{j+1} - v_j) F(v_j)(1 - F(v_j)) \\
&= \sum_{j=1}^{K-1} (v_{j+1} - v_j) P_j (1 - P_j), &(4)
\end{aligned}$$

where $P_j = \sum_{l=1}^j p_l$ is the cumulative probability of the first j categories. The second equality follows known properties of the CRPS, see e.g. Gneiting et al. (2007, Section 4.2). The expression at (4) is identical to the expression for the ERPS at (1) if $v_{j+1} - v_j = 1$ for all j, i.e., if all support points of X are exactly one unit apart. We will focus on this case in the following, and we investigate the properties of the CRPS entropy in this setup.

For concreteness, suppose the support of X is given by $v_1, v_2, \dots, v_K = 0, 1, \dots, K-1$. Furthermore, for a given integer $n \in \mathbb{N}$ and $s = 1, 2, \dots, n$, define the shifted random variables $X_s^n = X + s/(n+1)$, and $\mathbb{P}(X_s^n = j + s/(n+1)) = \mathbb{P}(X = j)$ for $j = 0, 1, \dots, K-1$. Figure 3.4 illustrates this construction for $n = 3$ (top panel) and $n = 20$ (bottom panel).

For given n and $\underline{p}$, we next consider the following family of mixture distributions:

$$\mathcal{F}_n(\underline{p}) = \left\{ Z : \mathbb{P}(Z \leq z) = \sum_{s=1}^{n} \omega_s^n F_s^n,\ 0 \leq \omega_s^n \leq 1, \sum_{s=1}^{n} \omega_s^n = 1. \right\}.$$

That is, $\mathcal{F}_n(\underline{p})$ collects all distributions that can be constructed as a finite mixture of the random variables $X_1^n, X_2^n, \dots, X_n^n$, where $F_s^n = F(z - s/(n+1))$ is the CDF of the s-th mixture component, and F is the CDF of the discrete random variable X with support points $0, 1, 2, \dots, K-1$ and associated probabilities $\underline{p}$. An interesting special case arises for $n \to \infty$ and $\omega_s^n = 1/n$ for all $s = 1, 2, \dots, n$, yielding a piecewise uniform distribution between $[0, 1], (1, 2], \dots, (K-1, K]$, as alluded to in the bottom panel of Figure 3.4. Such a setup is sometimes assumed for quantifying survey histograms (see e.g. Glas, 2020). Another practically relevant special case arises when n is odd, and $\omega_{(n+1)/2}^n = 1$, i.e., the 'central' mixture component receives a weight of one. In this case, we obtain the distribution assumed by the mass-at-midpoint method, with support at $0.5, 1.5, \dots, K-0.5$ and associated probabilities $\underline{p}$.

Consider a forecast histogram with bins $[0, 1], (1, 2], \dots, (K-1, K]$. Then for a given choice of n, each random variable in the family $\mathcal{F}_n(\underline{p})$ yields the same histogram, in the sense that $\mathbb{P}(Z \in \text{bin}_j) = p_j$ for each $Z \in \mathcal{F}_n(\underline{p})$, where bin_j denotes the interval that defines the jth histogram bin. All members of $\mathcal{F}_n(\underline{p})$ hence yield the same ERPS. Furthermore, Eq. (4) implies that all shifted random variables $X_s^n, s = 1, \dots, n$, yield the same expected CRPS. By contrast, the expected CRPS typically differs across members that are non-trivial mixtures of the X_s^n's, i.e., members that place strictly positive weight ω_s^n on at least two components s. The following result describes how the ERPS summarizes uncertainty across all members of $\mathcal{F}_n(\underline{p})$.

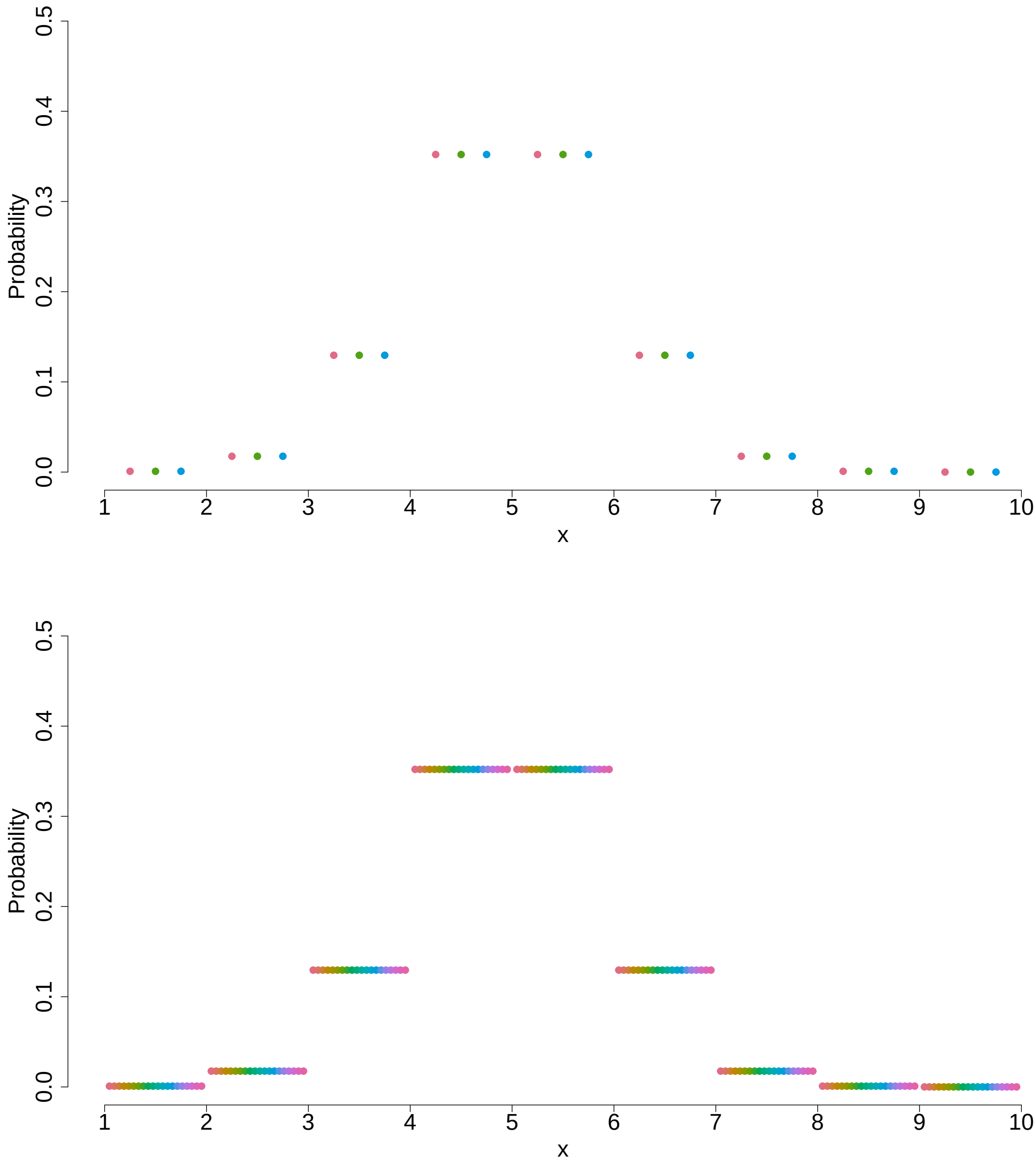


Figure 3.4: Illustration of the shifted random variables X_s^n, for $n = 3$ (top panel) and $n = 20$ (bottom panel). Each color corresponds to one value $s \in \{1, 2, \ldots, n\}$.

Proposition 1. *Consider the CDF $\sum_{s=1}^{n} \omega_s^n F_s^n$ of a member of the family $\mathcal{F}_n(\underline{p})$ described above, and let $ECRPS(\sum_{s=1}^{n} \omega_s^n F_s^n)$ denote the expected CRPS of this CDF. Then*

$$min_{\omega_1^n,\omega_2^n,\ldots,\omega_n^n \in \Delta_n} \; ECRPS\left(\sum_{s=1}^{n} \omega_s^n F_s^n\right) = \sum_{j=1}^{K-1} P_j(1-P_j),$$

where Δ_n is the set of all nonegative weights that sum to one. Furthermore, the minimum is attained by setting $(\omega_1^n, \ldots, \omega_n^n)$ equal to a unit vector of length n (with $\omega_s^n = 1$ for exactly one value $s = s^*$, and $\omega_s^n = 0$ for all other values of s.)

Proof. The result follows directly from the fact that ECRPS is a strictly concave function of its (CDF-valued) argument, which in turn results from the CRPS being a strictly proper scoring rule (see Gneiting and Raftery, 2007, end of Section 2.1). Note that Gneiting and Raftery define strictly proper scoring rules in positive orientation. Since we use them in negative orientation (such that smaller scores are better), their 'strictly convex' must be replaced by 'strictly concave' in our setting. Strict concavity of the ECRPS function means that

$$\begin{aligned}
\text{ECRPS}(\sum_{s=1}^{n} \omega_s^n F_s^n) \quad &\geq \quad \sum_{s=1}^{n} \omega_s^n \quad \underbrace{\text{ECRPS}(F_s^n)}_{=\sum_{j=1}^{K-1} P_j(1-P_j)} \\
&= \quad \sum_{j=1}^{K-1} P_j(1-P_j),
\end{aligned}$$

with equality if and only if $\omega_1^n, \ldots, \omega_n^n$ is a unit vector. □

In words, the proposition states that $\text{ERPS}(\underline{p}) = \sum_{j=1}^{K-1} P_j(1-P_j)$ is the minimal CRPS entropy in the family of distributions $\mathcal{F}_n(\underline{p})$, all of which are consistent with the same survey histogram $\underline{p}$. The analysis also implies that in the given setup of a histogram with bins $[0,1], (1,2], \ldots, (9,10]$, each shifted random variable X_s^n is compatible with $\underline{p}$, and its ECRPS coincides with the ERPS of the histogram. Hence, we can identify several numerical random variables replicating the ERPS' assessment of uncertainty.

The specific bin width of one considered here is not essential: Choosing equal-sized bins of another length c would render the ECRPS of each X_s^n equal to c times the ERPS of $\underline{p}$ (see Eq. 4). By contrast, using bins of different length would make the link between

ECRPS and ERPS less clear, which seems natural: Given that the ERPS is based on an ordinal interpretation of the bins, it cannot usefully reflect information on differences in bin length.

The family $\mathcal{F}_n(\underline{p})$ of distributions we consider ensures that each member is compatible with $\underline{p}$. Furthermore, it is easily possible to characterize the individual members' uncertainty (here, their ECRPS). This makes this family a suitable choice for studying the link between the ERPS and the uncertainty of distributions for numerical outcomes. At the same time, the result that the ERPS is the minimal expected CRPS does not hold across broader classes of distributions $\mathcal{G}(\underline{p})$ that are compatible with $\underline{p}$. We demonstrate this via the following example:

Example. Consider the bins $[0, 1], (1, 2], \ldots, (9, 10]$, and let $\underline{p} = \Big(0.1, \quad 0.1, \quad \ldots, \quad 0.1\Big)'$. The discrete distribution G which places probability 0.1 on the ten points $g_1, g_2, \ldots, g_{10} = 0.99, 1.99, 2.99, 3.99, 4.99,\ 5.01, 6.01, 7.01, 8.01, 9.01$ is compatible with $\underline{p}$, and attains an expected CRPS of 1.405 (see Eq. 4). This is strictly less than the ERPS of $\underline{p}$, which is given by 1.65.

3.8.4 Choice of Support Limits for Mass-at-midpoint (MAM) Method

Here we analyze the sensitivity of the MAM method to the choice of limit for the two outer bins, focusing on inflation expectations for brevity. Recall that the SCE's leftmost bin has an upper limit of -12, whereas the SCE's rightmost bin has a lower limit of $+12$. We consider the following three variants for closing the SCE's outer bins:

- **Narrow:** Leftmost bin equals $[-16, -12]$, rightmost bin equals $(12, 16]$. This implies a bin width of four, shared by the widest interior bins.

- **Medium:** Leftmost bin equals $[-20, -12]$, rightmost bin equals $(12, 20]$, i.e. doubling the bin width of the narrow variant.

- **Wide:** Leftmost bin equals $[-38, -12]$, rightmost bin equals $(12, 38]$. This choice corresponds to the maximal limit of 38 (or minimal limit of -38) that we impose in our implementation of the EMW method, following a proposal by Armantier et al (2017). As noted in Footnote 8, a wide choice of bin limits could also be motivated by the empirical occurrence of extreme *point* expectations of inflation.

The narrow and wide choices of bin limits are at the lower and upper end of what we would consider plausible. However, this assessment is necessarily subjective as no rigorous justification exists for one particular choice.

We implement the MAM method for these three choices of outer bins and consider the standard deviation obtained via the EMW method as a benchmark. Table 3.4 summarizes the empirical results, focusing on one-year-ahead inflation expectations (SCE, variable code Q9). For histograms that use only interior bins, the choice of outer bin limit is irrelevant by construction. Hence the differences in standard deviations are driven entirely by histograms that use at least one outer bin (about 37% of all histograms). The bottom panel of Table 3.4 thus presents summary statistics for this subsample.

For example, in the bottom panel of Table 3.4, the average standard deviation is 4.96 for the narrow choice, and 5.45 for the medium choice, corresponding to a relative increase of about 10%. The wide choice of outer bins generates substantially larger standard deviations, with an average of 7.89 (about 46% larger than the medium choice). Compared to EMW, all MAM variants yield higher mean and median values. However, the right tail of standard deviations tends to be higher for EMW than for the narrow and medium variants of MAM.

The practical relevance (or irrelevance) of these differences in standard deviations seems specific to the application considered. However, given the lack of a rigorous justification for the choice of bin limit for the MAM method, checking the robustness of empirical results to this parameter will often be necessary and burdensome.

3.8.5 Comparing Subjective and Objective Uncertainty

Here we provide a statistical justification for using the (expected and realized) RPS to compare subjective and objective measures of uncertainty and assess whether consumers' assessment of uncertainty is realistic. The latter aspect is economically relevant, in that misperceptions of uncertainty lead to economically suboptimal decisions in a wide range of situations (see e.g. Ben-David et al., 2013, and the references therein). Comparisons of expected and realized loss that are conceptually similar to the ones sketched here have been proposed by Clements (2014), Galvao and Mitchell (2019) and, Wei et al. (2017).

Table 3.4: Summary statistics for quantified standard deviations for the EMW and mass-at-midpoint (MAM) methods

Std. dev. of all histograms ($n = 103\,734$)						
Variant	5% quant.	25% quant.	Median	Mean	75% quant.	95% quant.
EMW	0.41	0.82	1.52	2.61	3.14	8.66
MAM-Narrow	0.00	1.02	2.04	2.71	3.63	8.27
MAM-Medium	0.00	1.02	2.06	2.90	3.98	8.97
MAM-Wide	0.00	1.02	2.10	3.81	5.69	12.42
Std. dev. of histograms using at least one outer bin ($n = 38\,737$)						
Variant	5% quant.	25% quant.	Median	Mean	75% quant.	95% quant.
EMW	1.37	2.48	3.89	4.99	6.82	11.77
MAM-Narrow	2.30	3.25	4.15	4.96	6.45	9.42
MAM-Medium	2.55	3.60	4.61	5.45	6.93	10.34
MAM-Wide	3.48	5.18	6.95	7.89	9.69	15.01

Note: Based on SCE histograms for inflation (one year ahead), between June 2013 and April 2020.

We consider a so-called prediction space setup (Gneiting and Ranjan, 2013) that models the joint distribution of expectations and realizations. We treat the K histogram probabilities $\underline{\mathbf{p}}$ as a random vector and denote the bin containing the realization by the discrete random variable $\mathbf{k}^* \in \{1, \ldots, K\}$. The sample space of interest, Ω, consists of forecast-observation pairs $(\underline{\mathbf{p}}, \mathbf{k}^*)$. We omit time indexes for simplicity; to obtain an intuition, subsequent realizations of $(\underline{\mathbf{p}}, \mathbf{k}^*)$ can be thought of as independent (whereas one would expect contemporaneous dependence between $\underline{\mathbf{p}}$ and $\mathbf{k}^*$, of course).[9] As in Ehm et al. (2016, Section 3.1), let $\mathbb{Q}$ be a probability measure on $(\mathcal{A}, \Omega)$, where $\mathcal{A}$ is a σ-field on Ω. The following result then provides a formal condition under which the expected and realized RPS coincide in expectation.

[9] It can be shown that the methodology of comparing ERPS to RPS remains valid under serial dependence in the forecast-observation tuples, as long as their joint process is strictly stationary. See Strähl and Ziegel 2017 for a technical treatment of a prediction space under serial dependence.

Assumption 1. *Assume that there is some information set $\mathcal{F} \subseteq \mathcal{A}$ such that*

$$\mathbb{Q}(\mathbf{k}^* = k|\mathcal{F}) = \mathbf{p}_k$$

holds almost surely for $k = 1, \ldots, K$, where $\mathbb{Q}(\mathbf{k}^* = k|\mathcal{F})$ is the true conditional probability that $\mathbf{k}^* = k$ (conditional on the information set $\mathcal{F}$), and $\mathbf{p}_k$ is the kth element of $\underline{\mathbf{p}}$.

Proposition 2. *Under Assumption 1, it holds that $\mathbb{E}(RPS(\underline{\mathbf{p}}, k^*)) = \mathbb{E}(ERPS(\underline{\mathbf{p}}))$.*

Proof. We have that

$$\begin{aligned} \mathbb{E}(\mathrm{RPS}(\underline{\mathbf{p}}, \mathbf{k}^*)) &= \mathbb{E}(\mathbb{E}(\mathrm{RPS}(\underline{\mathbf{p}}, \mathbf{k}^*)|\mathcal{F})) \\ &= \mathbb{E}(\sum_{k=1}^{K} \mathbf{p}_k \ \mathrm{RPS}(\underline{\mathbf{p}}, k)) \\ &= \mathbb{E}(\mathrm{ERPS}(\underline{\mathbf{p}})), \end{aligned}$$

where the first equality follows from the law of iterated expectations, the second equality follows from Assumption 1, and the final equality follows from the definition of ERPS. □

Assumption 1 requires that the probability forecast $\underline{\mathbf{p}}$ is correctly specified, in the sense that there is *some* information set relative to which the forecast is optimal. As noted by Gneiting and Resin (2022), the assumption is equivalent to $\underline{\mathbf{p}}$ being auto-calibrated (Tsyplakov, 2013), a notion of unbiasedness studied in the forecast evaluation literature. Under Assumption 1, Proposition 2 states that the RPS and ERPS of $\underline{\mathbf{p}}$ coincide in expectation. As a simple example (loosely following Gneiting et al., 2007, Table 1), let $Y = X + \varepsilon$, where both variables on the right are independently standard normal. Suppose for simplicity that there are only two outcome bins, $r_1 = (-\infty, 0]$ and $r_2 = (0, \infty)$. Consider forecaster A with $\mathbf{p}_1^{\mathrm{A}} = \Phi(-X), \mathbf{p}_2^{\mathrm{A}} = 1 - \Phi(-X) = \Phi(X)$. For forecaster A, Assumption 1 is satisfied with $\mathcal{F} = \sigma(X)$, the sigma algebra generated by X. In line with Proposition 2, it can be shown that the expected RPS and expected ERPS of forecaster A equal to 1/6. In the notation of Proposition 2, it holds that $\mathbb{E}(\mathrm{RPS}(\underline{\mathbf{p}}^{\mathrm{A}}, k^*)) = \mathbb{E}(\mathrm{ERPS}(\underline{\mathbf{p}}^{\mathrm{A}})) = 1/6$. For a second forecaster B with $\mathbf{p}_1^{\mathrm{B}} = \mathbf{p}_2^{\mathrm{B}} = 0.5$, Assumption 1 is satisfied with $\mathcal{F} = \emptyset$, the empty information set. The expected ERPS and expected RPS of forecaster B equal 1/4, confirming the intuition that B's forecast is less informative than A's forecast.

4 Assessing the Ex ante Uncertainty in the US SPF

4.1 Introduction

The insights of Bloom (2009) and the events during the Global Financial Crisis sparked a strong interest in the effects of changes in uncertainty on economic developments. The first challenge for corresponding investigations is the measurement of uncertainty. Therefore, many measures of uncertainty have been proposed in recent years, with those of Baker et al. (2016) and Jurado et al. (2015) being popular examples. In their comprehensive review article, Cascaldi-Garcia et al. (2021) distinguish between news-based (e.g. Baker et al., 2016), survey-based, econometric-based (e.g. Jurado et al., 2015), and market-based (e.g. the CBOE Volatility Index, VIX) measures.

Among the survey-based measures, Cascaldi-Garcia et al. (2021) further differentiate between ex ante and ex post measures. While ex post measures (e.g. Rossi and Sekhposyan, 2015) employ forecast errors and, thus, require survey forecasts and corresponding realizations, ex ante measures rely on survey forecasts only. If only point forecasts are provided, some measure of their cross-sectional dispersion among respondents is often used to assess uncertainty. This measure, of course, does not need to be informative about the uncertainty perceived by an individual respondent. The latter quantity can often be inferred from surveys that elicit probabilities about future developments. For instance, Altig et al. (2022) ask respondents for five-point distributions, where the five support points and their corresponding probabilities are chosen by the respondents. Asking for histogram forecasts is another widespread approach. In this case, the survey asks respondents to assign probabilities to - typically mutually exclusive - future events. If the forecast target is a continuous variable, these events are given by its realization in certain intervals. In general, these intervals cover the entire support of the random variable. Such intervals are used in the US Survey of Professional Forecasters (SPF), the SPF of the European Central Bank (ECB), but also in household surveys like the New York Fed's Survey of Consumer Expectations and the Deutsche Bundesbank Online Panel - Households. In this work, we will mainly focus on the US SPF's forecasts for inflation and growth.

When histogram forecasts are available, measures of average uncertainty can be constructed in different ways. As pointed out by Clements (2014), one can first calculate an average histogram across respondents and then gauge this histogram's dispersion, or one can first estimate the dispersion of each respondent's histogram and then average across these dispersions. We use the terminology by Clements et al. (2023) of referring to the latter measure as *aggregate uncertainty*, while the former measure corresponds to the *uncertainty of the aggregate histogram forecast*. If dispersion is measured by the variance, and if the true means and variances of the underlying distributions can be inferred from the histograms, it is well-known (see, e.g., Wallis, 2005) that the uncertainty of the aggregate histogram forecast equals the sum of the aggregate uncertainty and a term called *disagreement*, which is defined as the cross-sectional variance of the respondents' mean forecasts. While some studies focus on measuring the uncertainty of the aggregate histogram (see, e.g., Ganics et al., 2020), most investigate aggregate uncertainty (see, e.g., Zarnowitz and Lambros, 1987, Lahiri and Sheng, 2010, Boero et al., 2015 and Binder et al., 2022). We follow the latter approach for two reasons. First, we are going to document and adjust structural breaks which are related to the measurement of aggregate uncertainty. Second, we are going to propose a seasonal-adjustment approach for uncertainty based on the expected variance of a mean forecast. In the case of uncertainty of the aggregate histogram forecast, it is actually unclear what the corresponding mean forecast is.[1] Yet, it might be interesting to note that the share of aggregate uncertainty in the uncertainty of the aggregate histogram equals about two-thirds for the variables under study.[2]

[1] Apparently, it is often implicitly assumed that uncertainty of the aggregate histogram forecast corresponds to the expected variance of the combined mean forecast, i.e. of the average of the mean forecasts by all respondents. However, the expected variance of this mean forecast is equal to the aggregate uncertainty *minus* (not plus) expected disagreement under mild conditions, as shown in Proposition 1 of Knüppel and Krüger (2022).

[2] To be more precise, in our sample of the US SPF, the share of the average individual variance across all horizons considered equals 71% for growth and 65% for inflation, leaving shares of 29% and 35%,

The bin widths of the US SPF's histogram have undergone several changes since the start of the survey. For instance, the bin width for annual real GDP growth was reduced from two percentage points to one percentage point in 1992, most likely due to a lower variability of output observed since the mid-1980s. While most studies do not take these structural breaks into account, at least a few, relatively simple approaches have been proposed. Some authors simply focus on samples without breaks (e.g. Engelberg et al., 2009), but this is becoming increasingly difficult to justify because break-free samples are relatively short compared to the full sample now. Rich and Tracy (2010) suggest employing the definition with wider bins to the subsample(s) with narrower bins, such that each newly defined wide bin is assigned the sum of the probabilities of the corresponding narrower bins it contains.[3] This approach is based on the assumption that wider bins only imply a loss of information, but that each respondent always tries to map her subjective probability distribution to the histogram as precisely as possible, irrespective of the bin width. Based on a comparison between US SPF and ECB SPF respondents, Glas and Hartmann (2022) propose a drastically different assumption. They assume that each respondent assigns non-zero probabilities to a fixed number of bins only, irrespective of the bin width. A respondent ignoring the bin width completely implies that, for instance, doubling the bin width quadruples the variance of her histogram. To the best of our knowledge, Glas and Hartmann (2022) are the first to note that differences in bin widths can be associated with noticeable differences in measured uncertainty and that these differences require some form of adjustment. We are going to find that a change in bin widths indeed leads to a marked change in measured uncertainty, but that this change is less pronounced than suggested by Glas and Hartmann (2022).

The US SPF's histogram forecasts have a quarterly frequency. Since 1981q3, they are made for the current- and next-year variables. Thus, for each of these two fixed-event forecasts, there are four different forecast horizons, causing a seasonal pattern of the histograms' dispersion. While the dispersion is relatively large, for instance, for current-year forecasts made in the first quarter, it becomes small for current-year forecasts made

respectively, for disagreement. Details on our estimation approach for variance and disagreement can be found in Section 4.2.

[3] Of course, this approach only works if each endpoint of the wide bins coincides with an endpoint of the narrow bins.

in the fourth quarter, because much more information about the current year is available in the latter case. This seasonality hampers many types of analysis (see, for instance, Rossi and Sekhposyan, 2015, footnote 3). Therefore, we suggest a simple approach that combines the current- and next-year uncertainty forecasts of each quarter such that a fixed-horizon uncertainty forecast is obtained, which does not have a seasonal pattern. Our approach rests on the assumption that respondents make optimal forecasts based on a specific time series model and report the corresponding forecast error distribution.

The aim of our approach is related to Ganics et al. (2020), who combine current- and next-year density forecasts, which are derived from the histogram forecasts, to arrive at fixed-horizon density forecasts. From these fixed-horizon density forecasts, uncertainty measures can be calculated easily. However, density forecast combinations require the combination weights to be positive and to sum to one. In the case at hand, these restrictions can be problematic, as we will show below. Clark et al. (2022) apply cutting-edge econometric techniques to the US SPF's point forecasts, histogram forecasts, and point forecast errors to construct, among other quantities, an entire term structure of fixed-horizon uncertainty forecasts, providing an interesting measure of combined ex ante and ex post uncertainty. In contrast to our work, Ganics et al. (2020) and Clark et al. (2022) focus on the uncertainty of the aggregate histogram forecast.

In Section 4.2, we will describe the US SPF data and the construction of the uncertainty measure. The structural breaks will be investigated in Section 4.3. In Section 4.4, we propose and evaluate an approach for the seasonal adjustment of uncertainty forecasts. In Section 4.5, we use this approach for the seasonal adjustment of the US SPF's uncertainty measures. Comparisons to other uncertainty measures are performed in Section 4.6. Section 4.7 concludes.

4.2 US SPF Histograms and their Uncertainty

The US SPF is a quarterly survey released about two weeks after the first GDP release produced by the U.S. Bureau of Economic Analysis (BEA). Initially, the SPF was conducted by the American Statistical Association (ASA) together with the National Bureau of Economic Research (NBER), until the Philadelphia Fed took over in 1990q2

Table 4.1: Sample composition

variable	begin sample	end sample	min	average	max
		current year			
GDP deflator growth	1981q3	2022q1	7	32	51
real GDP growth	1981q3	2022q1	7	33	52
core CPI inflation	2007q1	2022q1	26	35	43
core PCE inflation	2007q1	2022q1	25	33	40
		next year			
GDP deflator growth	1981q3	2022q1	7	32	50
real GDP growth	1981q3	2022q1	7	33	52
core CPI inflation	2007q1	2022q1	26	34	43
core PCE inflation	2007q1	2022q1	25	32	39

Note: There is the possibility that an error occurred in the survey rounds 1985q1 and 1986q1, where the respondents were asked for their annual forecasts for the previous and current years instead. *Min, max* and *average* denote the minimum, maximum, and average number of forecasters per quarter over the sample period. The number of forecasters increased noticeably after the Federal Reserve Bank of Philadelphia took over the survey in 1990q2.

(see also Croushore and Stark, 2019, for a more detailed overview). The panelists, who usually work in forecasting firms, banks, and other financial institutions or academics, provide a large set of point predictions for, among many other variables, GDP deflator growth and real GDP growth.[4] For both variables and for core CPI and core PCE inflation, probabilistic forecasts are submitted as well. That is, the forecasters provide probabilities that the corresponding variable will realize within several pre-defined ranges (i.e. within certain bins of a histogram).

While some of the series date back to the late 1960s, the survey was continuously expanded and variables were added over the years. Generally, panelists issue forecasts for the current and next four quarters, as well as the current and next year, but also for longer horizons. The earliest observations in our sample date from 1981q3 when the forecast horizon 'next year' was first included for the variables of interest.

[4]For an exhaustive list of all variables and forecasts provided, see https://www.philadelphiafed.org/-/media/frbp/assets/surveys-and-data/survey-of-professional-forecasters/spf-documentation.pdf

Given a histogram for a continuous variable, uncertainty can be measured in different ways. Often, a smooth underlying probability distribution is assumed, and the parameters of this distribution are backed out from the histograms. Popular distributions for this purpose are given by the generalized Beta distribution (e.g. in Engelberg et al., 2009) and the normal distribution (e.g. in Giordani and Söderlind, 2003), for which dispersion measures are easy to calculate. Sometimes, instead of using a smooth distribution, it is assumed that the distribution is discrete with a point mass at the middle of each histogram bin (mass-at-midpoint approach, e.g. in Glas and Hartmann, 2022), which allows a straightforward calculation of the variance, for instance. Finally, an interesting distribution-free approach to measure uncertainty based on entropy is proposed by Rich and Tracy (2010) and Krüger and Pavlova (2021).

To compute the standard deviation, our preferred measure for uncertainty, from the individual histograms, we fit a distribution to the reported empirical probabilities, where the choice of the latter depends on the number of bins with strictly positive probability. For sparse histograms, i.e. the forecaster uses only one or two bins, we assume a symmetric triangular distribution following Engelberg et al. (2009). That is, in the case where all the probability mass is assigned to a single bin, the support of the triangle is equivalent to the range of the corresponding interval. We only consider two-bin histograms with probabilities assigned to adjacent bins. In the simplest case, where the forecaster uses two bins of equal width and equal probabilities, the support of the isosceles triangle covers the entirety of both bins. In all other instances, one can pin either the left or right endpoint of the support, depending on which interval is deemed more probable by the respondent. Additionally, in the two-bin case, we account for the possibility that the intervals have different widths, as described in greater detail in Appendix 4.8.1.

When the forecaster uses three or more intervals, we choose the normal distribution, because we are interested in first and second moments only, and the assumption of a normal distribution makes backing out parameters a robust procedure. We choose μ and σ such that the average squared difference between the empirical CDF and the CDF of the normal distribution is minimized. As starting parameters for the optimization, we use the mean and standard deviation computed via a simple mass-at-midpoint assumption. In all instances with positive probability allocated to open-ended intervals, following

Andrade et al. (2012) and Abel et al. (2016), we assume that the open interval has double the width of the adjacent closed interval.

This procedure is applied to all individual histograms from the US SPF for the variables GDP deflator growth, real GDP growth, core CPI, and core PCE inflation for two forecast horizons - current and next year. The sample for GDP deflator growth and real GDP growth covers the period 1981q3 to 2022q1, and the sample for core CPI and core PCE inflation the period 1981q3 to 2022q1. Table 4.1 provides an overview of the sample details.

These forecasts are *fixed-event* forecasts. That is, in each survey round within a calendar year, the probability assessments refer to the *annual-average over annual-average* percent change in the level of the GDP deflator for GDP deflator growth and of real GDP in the case of real GDP growth.[5] In either case, respondents provide forecasts for the current year, i.e. the year when the survey is conducted, and the subsequent year such that with each survey round within a given year, the forecast horizon declines. More formally, the annual-average over annual-average growth in the current and next year is defined as

$$\text{current-year growth rate: } \frac{\frac{1}{4}\left(p_1+p_2+p_3+p_4\right)}{\frac{1}{4}\left(p_{-3}+p_{-2}+p_{-1}+p_0\right)}-1 \tag{1}$$

$$\text{next-year growth rate: } \frac{\frac{1}{4}\left(p_5+p_6+p_7+p_8\right)}{\frac{1}{4}\left(p_1+p_2+p_3+p_4\right)}-1 \tag{2}$$

where the time index 1 refers to the first quarter of the current year, and p refers to the level of the variable. Often, one might be interested in a forecast of a growth rate for a *fixed horizon*, say the four-quarter-ahead quarter-on-quarter growth rate, such that the variable of interest would be given by

$$\text{four-quarter-ahead quarter-on-quarter growth rate: } \frac{p_{s+4}}{p_{s+3}}-1$$

where s denotes the current quarter. The corresponding four-quarter-ahead year-on-year growth rate is given by $p_{s+4}/p_s - 1$.

[5] For core CPI and core PCE inflation, the reported probabilities consider the *fourth-quarter over fourth-quarter* percentage change in the level of the core CPI (PCE) index. Obviously, the forecasts for these growth rates are fixed-event forecasts as well.

4.3 Structural Breaks

The resulting aggregate uncertainty, measured by the average standard deviation across all respondents in each survey round, is shown in Figure 4.1. Obviously, there are four large changes in uncertainty. For GDP deflator growth (henceforth often simply *inflation*), the uncertainty dropped markedly in 1992q1 and 2014q1. For real GDP growth (henceforth often simply *growth*), the uncertainty dropped noticeably in 1992q1 and increased strongly in 2020q2. All these dates coincide with changes in the bin width of the US SPF's histograms.

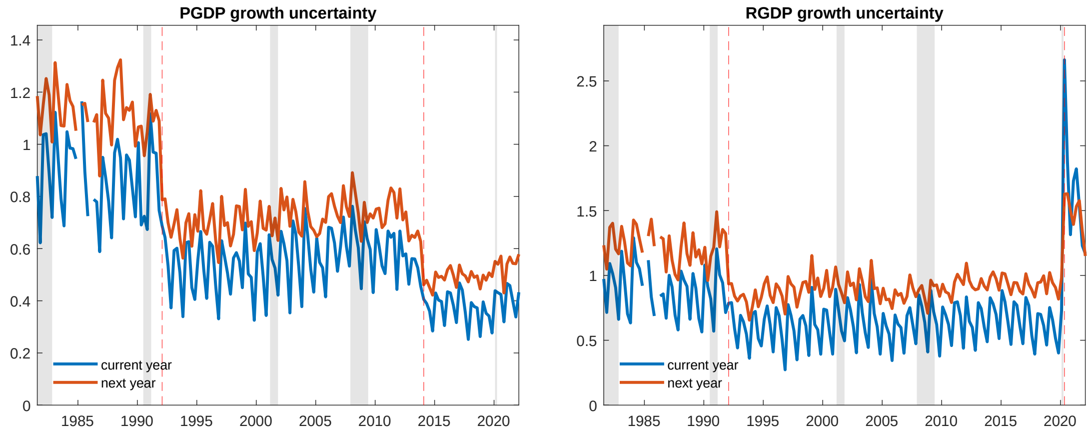


Figure 4.1: Aggregate uncertainty of GDP deflator growth (left panel) and real GDP growth (right panel) concerning the current and the next year. Uncertainty is measured by the standard deviation and averaged over forecasts for each survey round. The red vertical lines indicate dates with changes in bin widths. Shaded areas indicate recessions as dated by the NBER.

In 1992q1, the bin width for inflation and growth declined from 2 percentage points (pp) to 1 pp. In 2014q1, the bin width for inflation halved again, reaching 0.5 pp. In 2020q2, the bin widths for growth became larger and distinct. Only the central bin remained at a width of 1 pp, while the adjacent bins had 1.5 pp. The bin widths further increased with their distance from the central bin, arriving at 6 pp for the widest of the closed bins. Figure 4.2 shows how the bin widths and also the number of bins changed over time.

Figure 4.2: Current-year GDP deflator (left panel) and real GDP (right panel) growth forecasts together with histogram bins over time. Black dashed lines mark periods with changes in the bin definitions. The graph is based on presentation slides for study of Clark et al. (2022).

The halving of bin widths for inflation and growth in 1992q1 was most likely related to the reduced volatility of these variables after the onset of the Great Moderation in the mid-1980s (cf. McConnell and Perez-Quiros, 2000). The increase in the bin widths for growth in 2020q2 was supposed to accommodate the large swings in growth due to the starts and ends of lockdowns during the COVID-19 pandemic. The second halving of bin widths for inflation in 2014q1 was probably again due to the low volatility of this variable in previous years.

To the best of our knowledge, the literature that investigates the impact of histogram features on the forecast uncertainty provided by respondents is still very limited. A study by Becker et al. (2023) considers the behavior of consumers when confronted with histograms, which differ concerning their coarseness. In one of their randomized controlled trials (RCT), each group of consumers gets a histogram that differs from a baseline scenario concerning the number of bins, while the set of closed bins always covers the same interval. For instance, there can be only 4 (half-)closed histogram bins ([-12,-8),[-8,0),[0,8),[8,12]) covering the interval [-12,12], or there can be 12 (half-)closed bins on the same interval. This variation is called ‘Centralization’. In a different treatment, the number of bins remains constant, while the width of the bins and, thus, the interval covered by the union of all closed bins varies. For instance, while there are always 8 bins, these bins can cover the interval [-48,48] or the interval [-3,3]. In the former case, the bins have widths ranging from 8 to 16 pp, whereas in the latter case, the widths range from 0.5 to 1 pp. This treatment is labeled 'Compression'.

Becker et al. (2023) show that histogram features have noticeable effects on measured forecast uncertainty. In the 'Centralization' treatment, fewer bins and, thus, larger bin widths lead to an increase in measured uncertainty. In the 'Compression' treatment, a larger interval covered by the union of all closed bins and, thus, having larger bin widths also amplifies measured uncertainty in their survey. Taken together, this evidence suggests that a decrease in bin width is associated with a decrease in measured uncertainty. However, the respondents in the RCT are consumers, and hence it is unclear whether professional forecasters would behave similarly under equivalent conditions.

In what follows, we will try to assess whether and how a change in bin widths might affect the measured uncertainty of professional forecasters. Based on the available data, we will not be able to investigate if a change in the number of bins also has an impact on measured uncertainty. Therefore, we have to maintain the assumption that a change in the number of bins, ceteris paribus, leaves measured uncertainty unaffected for professional forecasters.

As mentioned above, since 2007q1, the US SPF also asks for histogram forecasts for two other inflation measures, namely for core PCE inflation and core CPI inflation. As shown in Figure 4.3, the uncertainty concerning these inflation measures appears to be relatively stable around 2014q1. Moreover, the evolution of annual inflation rates as displayed in Figure 4.4 does not indicate any major differences in the volatility of GDP deflator growth in the years before and after 2014, and the same holds for core CPI inflation and core PCE inflation.[6] This claim would become clearly invalid if the (symmetric) window around 2014 became large enough to include the year 2021, but seems to be justified for smaller window sizes.

[6]Note that annual core CPI inflation is measured by the year-on-year growth rates of the fourth-quarter averages of the core CPI level. The fourth-quarter average CPI level is the average of the corresponding three monthly levels.

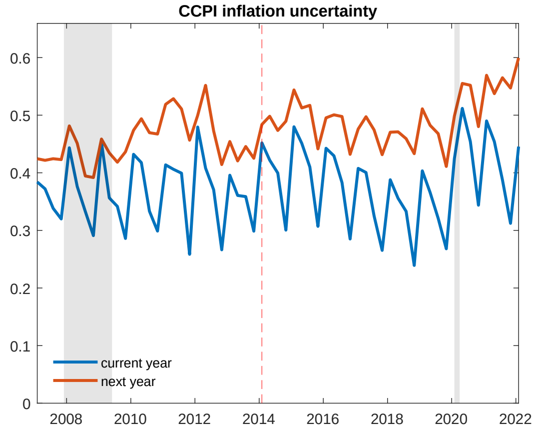


Figure 4.3: Aggregate uncertainty of core CPI inflation (left panel) and core PCE inflation (right panel) concerning the current and the next year. Uncertainty is measured by the standard deviation and averaged over forecasts for each survey round. The red vertical line indicates the dates with changes in bin widths for GDP deflator growth. Shaded areas indicate recessions as dated by the NBER.

In the spirit of McConnell and Perez-Quiros (2000), we estimate a first-order autoregressive model for annual GDP deflator growth and regress the absolute values of the resulting residuals from the period 2009 to 2018 on a constant and a dummy variable which equals 0 before 2014 and 1 afterward.[7] We repeat the same exercise with quarterly GDP deflator growth. As shown in Table 4.2, in both cases, the coefficient for the dummy variable is not significantly different from zero, supporting the hypothesis that a structural change in GDP deflator growth uncertainty in 2014q1 is not likely.[8]

For core CPI inflation and core PCE inflation, we can directly test the hypothesis of no change in uncertainty based on the histogram forecasts. This seems clearly preferable to the regression-based approach used for GDP deflator growth because we are interested in the ex ante uncertainty as perceived by respondents. While it is not impossible that this uncertainty is related to the magnitude of shocks that occur to the variable under study, it is not evident that respondents can predict these magnitudes or that they expect recently observed magnitudes to persist over the forecast horizon.

[7]The reason for choosing a window of ten years will be discussed below.

[8]It might be interesting to note that the same tests as described in Table 4.2, if used in a ten-year window around 1992 (i.e., the date of the first change in bin widths), would reject the null hypothesis of no change in uncertainty at the 10% level with annual data and at the 5% level with quarterly data.

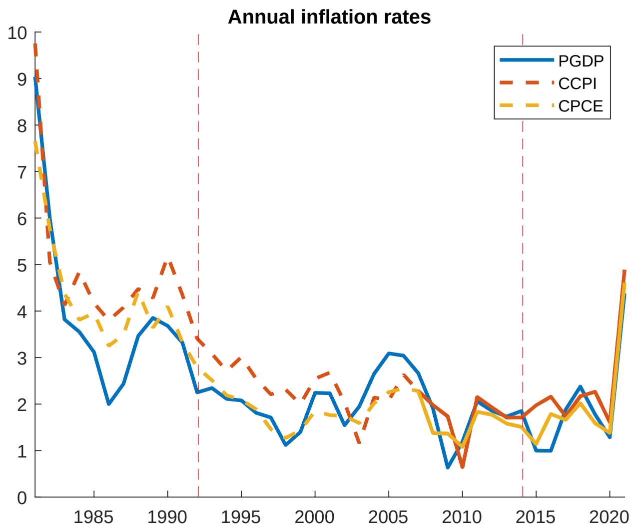


Figure 4.4: Annual inflation rates based on the GDP deflator (PGDP), core CPI (CCPI), and core PCE (CPCE) index from 1981 until 2021. Solid lines indicate the availability of corresponding histogram forecasts in the US SPF. The definitions of the annual inflation rates are identical to those used for the histogram forecasts, i.e., these are year-on-year rates of the fourth-quarter levels for the core CPI and the core PCE index and year-on-year rates of annual levels for the GDP deflator. Further details are provided in the text. The red vertical lines indicate dates with changes in bin widths for GDP deflator growth in the US SPF.

Table 4.2: Test for change in volatility of GDP deflator growth around 2014

regression: $\lvert\varepsilon_t\rvert = c_\varepsilon + \beta\, D_t + u_t$					
annual (2009 to 2018)			quarterly (2009q1 to 2018q4)		
β	standard error	p-value	β	standard error	p-value
-0.04	0.28	0.90	0.29	0.18	0.11

Note: ε_t is the residual of the regression equation $y_t = \rho\, y_{t-1} + \varepsilon_t$ estimated for the sample 1981 to 2021 (quarterly data: 1981q3 to 2022q1). The dummy variable D_t equals 0 for $t < 2014$ (quarterly data: $t < 2014q1$) and 1 for $t \geq 2014$ (quarterly data: $t \geq 2014q1$). The p-values result from standard t-tests of $H_0 : \beta = 0$. Both equations are estimated by OLS.

Denoting the inflation uncertainty forecast for inflation measure π made in year t and quarter i for the year $t+j$ by $\sigma^{\pi}_{t+j|t,i}$, we estimate the 16 equations given by

$$\sigma^{\pi}_{t+j|t,i} = c_{\pi,i,j} + \beta_{\pi,i,j}\, D_t + \varepsilon_{t,\pi,i,j} \tag{3}$$

with $\pi \in \{\text{core CPI inflation, core PCE inflation}\}$, $i = 1,2,3,4$ denoting the quarter the forecast is made, $j = 0,1$ denoting the forecast horizon in years (i.e. 0 for the current year, 1 for the next year), $t = t_{min}+1,\ldots,t_{max}$ denoting the year the forecast is made, and ε_t denoting an error term. The 16 equations result from all possible combinations of π, i, and j. The shortest forecast horizon corresponds to $i = 4, j = 0$, which is the forecast for the current year made in the fourth quarter. The longest forecast horizon corresponds to $i = 1, j = 1$, which is the forecast for the next year made in the first quarter. The dummy variable D_t is given by

$$D_t = \begin{cases} 0 & \text{if } t < 2014 \\ 1 & \text{if } t \geq 2014. \end{cases} \tag{4}$$

We test the null hypothesis $H_0 : \beta_{\pi,i,j} = 0$ for all π, i, j jointly, i.e. for all 16 equations jointly. By varying the size of the symmetric estimation window around the date 2014, we can find the sample with the largest p-value for H_0. Below, we will use this sample to gauge the effect of a bin width change on the uncertainty measured for GDP deflator growth. It turns out that the ten-year window from 2009 to 2018 gives the largest p-value and does not provide evidence against the absence of structural change. Note that this is not due to power problems, because the null hypothesis of no structural change is rejected for smaller as well as for larger samples.

In summary, we do not find any empirical evidence for a structural break in true inflation forecast uncertainty in 2014. Therefore, we proceed under the assumption that this uncertainty did not change in 2014.

Consequently, in what follows, we are going to assume that a change in the bin width of a forecast histogram can have an effect on the uncertainty measured based on this histogram. Specifically, we assume that the measured standard deviation is related to the bin width by

Table 4.3: Tests for change in forecast uncertainty in 2014 for core CPI and core PCE inflation

t_{min}	2011	2010	2009	2008	2007
t_{max}	2016	2017	2018	2019	2020
p-value	0.02	0.08	0.57	0.23	0.00

Note: p-values for F-test of $H_0 : \beta_{\pi,i,j} = 0$ for all π, i, j jointly in regression equation (3), which is estimated by OLS.

$$\sigma_{meas} = c\, x^{\gamma} \tag{5}$$

where c is some constant and x denotes the bin width, i.e. the coarseness of the histogram. In the cases studied here, x denotes pp per bin. Changes in x then change measured uncertainty according to

$$\Delta \ln (\sigma_{meas}) = \gamma\, \Delta \ln (x)\,. \tag{6}$$

We choose this specification because it relies on a single parameter, which can be determined even if only a single bin width change is observed. Moreover, the specification nests two important special cases, namely $\gamma = 0$ and $\gamma = 1$. The literature, in general, seems to assume that $\gamma = 0$, implying that the bin width does not affect measured uncertainty. As we have argued above, this assumption does not seem to hold. Glas and Hartmann (2022) assume that $\gamma = 1$, arguing that respondents might always assign positive probabilities to a fixed number of bins only, which implies that they ignore the bin widths when making their histogram forecasts.[9] Under this assumption, one could imagine, for instance, that a forecaster first assigns positive probabilities to three bins (20%, 65%, and 15%, say) of unknown width, then looks for the bin containing her point forecast and assigns the probabilities to this bin and the adjacent bins.

We determine γ based on the change in measured uncertainty for GDP deflator growth in 2014q1. We use a regression similar to equation (3), namely

[9] Glas and Hartmann (2022) actually use the variance to measure uncertainty. If we square both sides of (5), obviously we get $\sigma^2_{meas} = c^2\, x^{2\gamma}$, implying that doubling the bin width leads to a fourfold increase in the measured variance if $\gamma = 1$.

$$\ln\left(\sigma^{\pi}_{t+j|t,i}\right) = c_{i,j} + \gamma_{i,j}\,\ln\left(x^{\pi}_{t,i}\right) + \varepsilon_{t,i,j} \tag{7}$$

with π here denoting GDP deflator growth, and the variable $x^{\pi}_{t,i}$ is a measure of the coarseness of the histogram. Here, $x^{\pi}_{t,i}$ is given by

$$x^{\pi}_{t,i} = \begin{cases} 1 & \text{if } t < 2014 \\ 0.5 & \text{if } t \geq 2014, \end{cases} \tag{8}$$

which corresponds to the histogram bin widths in pp. We set $t_{min} = 2009$ and $t_{max} = 2018$ because of the results reported in Table 4.3.[10] Hence, we simply try to explain the measured uncertainty of GDP deflator growth by a constant and a dummy-type variable capturing the effect of the change in the bin width. The estimation results are displayed in Table 4.4.

Table 4.4: Effects of bin width change on GDP deflator growth uncertainty in 2014

i	j	$c_{\pi,i,j}$	std. err.	$\gamma_{i,j}$	std. err.
4	0	−0.779	0.036	0.698	0.073
3	0	−0.549	0.032	0.629	0.065
2	0	−0.504	0.029	0.567	0.058
1	0	−0.427	0.035	0.617	0.071
4	1	−0.392	0.027	0.576	0.055
3	1	−0.321	0.028	0.577	0.058
2	1	−0.321	0.033	0.540	0.068
1	1	−0.279	0.034	0.577	0.069

Note: Regression results of equation (7). i corresponds to the quarter of the current year when the forecast is made, j is the forecast horizon in years. The forecast horizon thus increases by one quarter with each row. The estimation sample is 2009q1 to 2018q4.

[10] Based on our assumption that the open bins have twice the width of the closed bins, the average bin widths used by respondents are marginally higher than 1 and 0.5. However, note that for the estimation of γ, only the ratio between $x^{\pi}_{t,i}$ before and after 2014 matters, such that the width of the open bins in the definition of $x^{\pi}_{t,i}$ can be ignored.

Obviously, $\gamma_{i,j}$ equals about 0.6 for most forecast horizons. Therefore, we test the joint null hypothesis $H_0 : \gamma_{i,j} = \gamma_{k,l}$ for all $(i, j) \neq (k, l)$, i.e. we test whether all coefficients $\gamma_{i,j}$ are equal. The Wald test yields a p-value of 0.84. This result implies that there is no evidence against a bin width change having the same effect for all forecast horizons. Based on this useful insight, we proceed by estimating the restricted version of regression (7) given by

$$\ln\left(\sigma^{\pi}_{t+j|t,i}\right) = c_{\pi,i,j} + \gamma \ln\left(x^{\pi}_{t,i}\right) + \varepsilon_t. \tag{9}$$

Table 4.5: Average effect of bin width change on GDP deflator growth uncertainty

i	j	$c_{\pi,i,j}$	std. err.	γ	std. err.
4	0	−0.814	0.023		
3	0	−0.560	0.023		
2	0	−0.494	0.023		
1	0	−0.433	0.023		
4	1	−0.384	0.023	0.598	0.023
3	1	−0.314	0.023		
2	1	−0.301	0.023		
1	1	−0.272	0.023		

Note: Regression results of the restricted equation (9). i corresponds to the quarter of the current year when the forecast is made, j is the forecast horizon in years. The forecast horizon thus increases by one quarter with each line. Note that the standard errors are identical for each $c_{\pi,i,j}$ by construction of the regression equation. The standard error of γ differs slightly, but, by coincidence, turns out to be identical when rounding to the third decimal place. The estimation sample is 2009q1 to 2018q4.

The estimation results in Table 4.5 show that γ equals about 0.6. Looking at the standard error of the estimate of γ, one can clearly reject both assumptions encountered in the literature concerning the effect of a change in the histogram bin width on measured uncertainty. The usual assumption of $\gamma = 0$ has a p-value of virtually zero, implying that measured uncertainty does change in response to changes in bin width. The assumption $\gamma = 1$ of Glas and Hartmann (2022) has a p-value of virtually zero as well, implying that the standard deviation does not change one-to-one with the bin width. Our result shows that an intermediate assumption is warranted.[11] With $\gamma = 0.6$, halving the bin width yields a measured standard deviation that is smaller by a factor of 0.66, i.e. it causes the measured standard deviation to decrease by about one-third. This result simply follows from $\gamma\ \ln(1/2) = -0.416$ and $\exp(-0.416) = 0.66$.[12]

Denoting the estimated value of γ by $\hat{\gamma}$, we can determine the break-adjusted GDP deflator growth uncertainty by setting

$$\tilde{\sigma}^{\pi}_{t+j|t,i} = \sigma^{\pi}_{t+j|t,i}\ \left(x^{\pi}_{t,i}\right)^{\hat{\gamma}}. \tag{10}$$

Since $x^{\pi}_{t,i} = 1$ from 1992q1 to 2013q4, the break-adjusted uncertainty coincides with the original uncertainty over this period. Before 1992q1, the bin width is equal to 2 pp, such that $x^{\pi}_{t,i}$ equals 2 from the start of the sample until 1991q4.

This type of break adjustment, however, cannot directly be applied to real GDP growth uncertainty since 2020q2. This is due to the fact that from that date onward, many different bin widths co-exist, and it is unclear which value for $x^{y}_{t,i}$, i.e. for the coarseness of the output histograms would be appropriate. Therefore, we gauge the value for $x^{y}_{t,i}$ by applying the bin definitions prevailing before 2020q2 to the histogram forecasts since 2020q2 and determining the corresponding uncertainty. For instance, if a respondent assigned 60% probability to the fourth bin (width of 3 pp since 2020q2) and 40% probability to the fifth bin (width of 1.5 pp since 2020q2), we simply assume a width of 1 pp for each bin (the width of all closed bins in 2020q1) to calculate the standard deviation. This approach is feasible because the number of bins did not change in 2020q2.

[11] One could imagine, for instance, that, in contrast to the assumption of Glas and Hartmann (2022), the number of bins that are assigned positive probabilities by a respondent increases somewhat when the bins become smaller. That is a respondent who used to assign three positive probabilities before the width change might assign four or five positive probabilities thereafter.

[12] Interestingly, Becker et al. (2023) find a similar result in one of their 'Compression' cases, which most closely resembles our case. The measured standard deviation in their baseline equals 3.08 when fitting beta distributions. When the bin widths are halved, the measured standard deviation equals 1.84, which would translate into $\gamma = \frac{\ln(1.84/3.08)}{\ln(1/2)} = 0.74$. However, their setting is different from ours and the US SPF in several respects, and in other, similar 'Compression' cases they find results implying other values for γ as well. Yet, their results clearly do not support the assumption $\gamma = 0$.

Denoting the aggregate real GDP growth uncertainty obtained in this way for each quarter since 2020q2 by $\acute{\sigma}^{y}_{t+j|t,i}$, we can exploit the relation

$$\frac{\sigma^{y}_{t+j|t,i}}{\acute{\sigma}^{y}_{t+j|t,i}} = \frac{x^{y}_{t,i,j}}{x^{y}_{2020,1,j}} \tag{11}$$

for $t = 2020, 2021, \ldots$ and $i = 2, 3, 4$ if $t = 2020$ and $i = 1, 2, 3, 4$ if $t > 2020$. The histogram coarseness $x^{y}_{t,i,j}$ equals 1 in 2020q1, corresponding again to the width of the closed bins in that quarter. Therefore, the values of $x^{y}_{t,i,j}$ since 2020q2 can simply be determined by the left-hand side of equation (11). Note that the way $\acute{\sigma}^{y}_{t+j|t,i}$ is calculated corresponds to the assumption $\gamma = 1$ of Glas and Hartmann (2022). If respondents essentially ignore the bin widths, and only pay attention to the order of the bins (i.e. whether a bin is, say, the central bin, the first bin with larger values than the central bin, etc.), the expected measured uncertainty for any hypothetical bin definition (having the same number of bins as in the original setting) can be found simply by applying this bin definition to the probabilities provided by the respondents. Correspondingly, (11) follows from (6) with $\gamma = 1$.

Our previous estimate $\hat{\gamma} \approx 0.6$ is obtained from a sample where the bin width was halved for all bins. One might argue that the change for real GDP growth in 2020q2 was fundamentally different since, for instance, the width of the central bin did not change (1 pp), while the width of the most extreme closed bins strongly increased (from 1 pp to 6 pp). However, we proceed by regarding $\hat{\gamma} \approx 0.6$ as a value which applies to any redefinition of the bins, and by regarding equation (11) as a general approach to determining the histogram's coarseness $x_{t,i}$ in every situation where bin widths change, if the number of bins remains constant.[13] We do so because both extreme assumptions that have been used in the literature so far, i.e. $\gamma = 0$ and $\gamma = 1$, are clearly rejected by the data on the break in GDP deflator growth in 2014q1, and because there is yet no guidance on the question whether an intermediate value of γ other than 0.6 might

[13]For the general approach, one would simply replace $x^{y}_{2020,1,j}$ by some $x_{t,i,j}$ equal to 1 for some year t and quarter i, where this period then serves as the base period for which the measured uncertainty is taken at face value. Note that we also could have used this approach for GDP deflator growth. The number of bins did not change in 2014q1, such that we could have calculated $x^{\pi}_{t,i,j}$ for $t \geq 2014$ by using the bin definitions from 2013q4 in equation (11). The resulting series of $x^{\pi}_{t,i,j}$ would simply equal 0.5.

be more appropriate in the situation encountered for real GDP growth. In our view, it would be very interesting to conduct a corresponding survey experiment, but we are not aware of such a study.[14]

We thus proceed by using equation (10) to obtain a break-adjusted uncertainty series for real GDP growth as well, i.e. we calculate the break-adjusted series $\tilde{\sigma}^y_{t+j|t,i}$ as

$$\tilde{\sigma}^y_{t+j|t,i} = \sigma^y_{t+j|t,i} \left(x^y_{t,i,j} \right)^{\hat{\gamma}} \tag{12}$$

with $\hat{\gamma}$ as in (10). Similar to GDP deflator growth, we have $x^y_{t,i} = 1$ from 1992q1 to 2020q1, such that the break-adjusted uncertainty coincides with the original uncertainty over this period. Again, $x^y_{t,i}$ equals 2 from the start of the sample until 1991q4. From 2020q2 onward, $x^y_{t,i,j}$ changes each period and depends on the forecast target j.

Figure 4.5 displays the resulting coarseness series. For GDP deflator growth, the histograms for the current and the next year always have identical coarseness, which corresponds to the width of the closed intervals. For real GDP growth, marked differences appear with the last bin width change. In comparison to the next-year forecasts, the current-year forecasts tend to have much more probability mass in less central bins. Since less central bins are wider, the ratio of the standard deviations in (11) is larger for current-year forecasts. For next-year forecasts, the coarseness measure also increases with the bin width change in 2020q2, but it remains smaller than before 1992q1.

Using the coarseness series in equations (10) and (12) finally gives the break-adjusted series, which are displayed in Figure 4.6. While the structural break in measured uncertainty of inflation in 2014q1 disappears by construction, the lack of visible structural breaks in 1992q1 for inflation and growth is remarkable. For inflation, uncertainly appears at best marginally larger before 1992q1 than in the years afterward. For growth, uncertainty seems to remain relatively constant from the start of the sample until about the late 1990s, when uncertainty tends to increase slightly. With the start of the Corona pandemic in 2020q2, growth uncertainty reaches higher levels than observed before, but

[14] Becker et al. (2023) is the only study we know, which investigates the effect of bin definitions on measured uncertainty. However, the authors always use a bin definition where bins actually become smaller with their distance from the central bin.

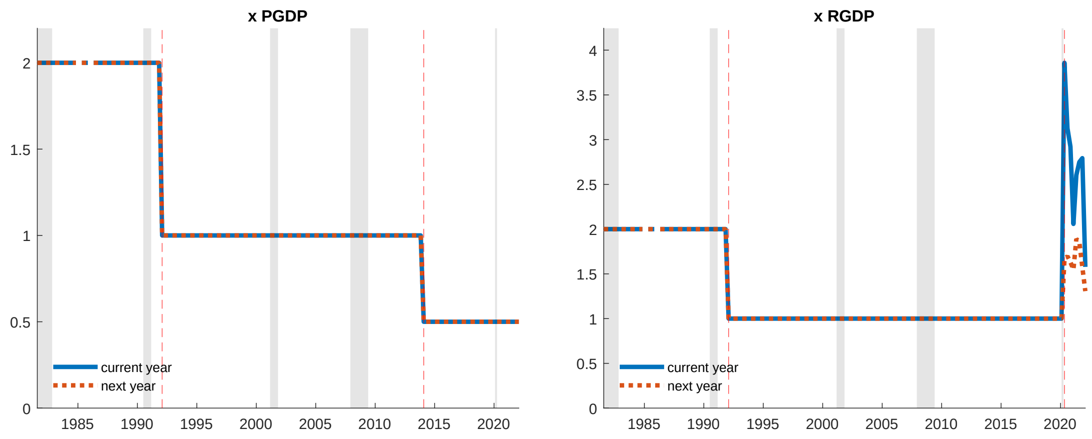


Figure 4.5: Coarseness of histograms as measured by $x^{\pi}_{t,i}$ for GDP deflator growth (left panel) and $x^{y}_{t,i,j}$ for real GDP growth (right panel). Red vertical lines indicate the dates with changes in bin widths. Shaded areas indicate recessions as dated by the NBER.

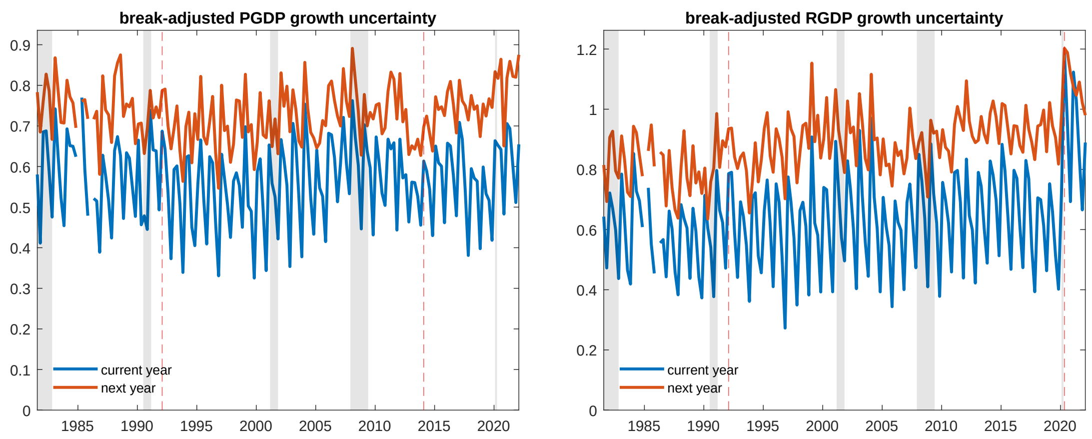


Figure 4.6: Break-adjusted uncertainty of GDP deflator growth (left panel) and of real GDP growth (right panel). Red vertical lines indicate the dates with changes in bin widths. Shaded areas indicate recessions as dated by the NBER.

the increase is considerably less drastic than conveyed by the unadjusted uncertainty series in Figure 4.1.

A notable feature that becomes more evident in the break-adjusted uncertainty series, although it is also present in the unadjusted series, is the lacking relationship between business cycles and uncertainty. There is a large literature investigating if rising uncertainty, especially concerning economic activity, causes recessions or if recessions increase uncertainty or both, but the link between high uncertainty and recessions is mostly regarded as a fact (see, for instance, Ludvigson et al., 2021). For growth, only the pandemic recession is clearly associated with higher uncertainty. For inflation, uncertainty appears to be relatively large around the recession related to the financial crisis of 2007-08. However, for all recessions except the pandemic one, it would be very difficult to spot them based on the uncertainty series in Figure 4.6. The following seasonal adjustment of the uncertainty series will corroborate this finding.

4.4 Seasonal Adjustment

4.4.1 A Seasonal-adjustment Approach for Fixed-event Uncertainty Forecasts

Due to fixed-event forecasting, the US SPF's uncertainty forecasts contain a seasonal pattern. We are going to investigate the uncertainty of optimal forecasts for the type of fixed-event forecasts under study, and we are going to seasonally adjust the uncertainty series based on those insights. The ECB SPF features the same type of fixed-event forecasts but has additional fixed-horizon forecasts. Therefore, we are going to evaluate our seasonal-adjustment approach using the ECB SPF before applying it to the US SPF.

As mentioned above, the forecasted variables differ in their transformation of the original (often higher-frequency) data in levels (growth rates of annual averages, year-on-year growth rates of fourth-quarter levels, etc). As shown by Patton and Timmermann (2011), these transformations can be approximated by

$$G_\tau \approx \sum_{s=s_1}^{s_2} \omega_{s-s_1+1} \times g_s,$$

where g_s is a high-frequency (here: month-on-month) growth rate with Wold representation

$$g_s = c + \sum_{j=0}^{\infty} \theta_j \times \varepsilon_{s-j}, \tag{13}$$

and ω_{s-s_1+1} is a known transformation-dependent weight applied to g in period s, and s_1, s_2 also depend on the transformation. G_τ is the value of G in period τ, where τ can have an annual, quarterly or monthly frequency in our application, the θ_i's are parameters with $\theta_0 = 1$, and ε_s is a white noise process with $E[\varepsilon_s^2] = \sigma_s^2$.

Assuming optimal forecasts, the error of a forecast for G_τ made in period $\tilde{s}$ is given by

$$\hat{\varepsilon}_{G_\tau|\tilde{s}} = \boldsymbol{\omega}' \, \boldsymbol{\Theta} \, \boldsymbol{\varepsilon}_{\tilde{s}} \tag{14}$$

with $\boldsymbol{\omega}' = \begin{bmatrix} \omega_m & \omega_{m-1} & \cdots & \omega_1 \end{bmatrix}$, $\boldsymbol{\Theta}$ being an $m \times m$ matrix of θ_j's (and zeros), $\boldsymbol{\varepsilon}_{\tilde{s}} = \begin{bmatrix} \varepsilon_m & \varepsilon_{m-1} & \cdots & \varepsilon_{\tilde{s}+1} & 0 & \cdots & 0 \end{bmatrix}'$, $m = s_2 - s_1 + 1$. The expected squared error of the forecast for G_τ is

$$E\left[\hat{\varepsilon}^2_{G_\tau|\tilde{s}}\right] = \boldsymbol{\omega}' \, \boldsymbol{\Theta} \, E\left[\boldsymbol{\Omega} \,| I_{\tilde{s}}\right] \, \boldsymbol{\Theta}' \, \boldsymbol{\omega} \tag{15}$$

with $E\left[\boldsymbol{\Omega} \,| I_{\tilde{s}}\right]$ being a matrix with diagonal

$$diag\left(E\left[\boldsymbol{\Omega} \,| I_{\tilde{s}}\right]\right) = \begin{bmatrix} \hat{\sigma}^2_{m|\tilde{s}} & \hat{\sigma}^2_{m-1|\tilde{s}} & \cdots & \hat{\sigma}^2_{\tilde{s}+1|\tilde{s}} & 0 & \cdots & 0 \end{bmatrix}$$

with $\hat{\sigma}^2_{\tilde{s}+h|\tilde{s}} = E\left[\varepsilon^2_{\tilde{s}+h} \,| I_{\tilde{s}}\right]$ and all off-diagonal elements equal to zero.

Concerning $\hat{\sigma}^2_{\tilde{s}+h|\tilde{s}}$, we simply assume that

$$\hat{\sigma}^2_{\tilde{s}+h|\tilde{s}} = \sigma^2_{\tilde{s}}, \tag{16}$$

implying a random-walk behavior of the variance. Many studies find that a random-walk specification for the log of the variance yields good results for macroeconomic data at least before the pandemic recession (see, for instance, Clark and Ravazzolo, 2015).[15] Under

[15] While the latter specification for the logarithm implies an increase in the expected variance itself due to Jensen's inequality, assumption (16) is a good approximation if the error term in the log-variance equation is not too large.

this assumption, the expected squared error of any forecast for a linear transformation of g_s can be determined if the parameters θ_j of equation (13) are known, and if at least one expected squared forecast error is available. The latter allows backing out $\sigma^2_{\tilde{s}}$ if the θ_j's are known.

In our case, *two* expected squared forecast errors are available, measured by the variance of the histograms for the current-year and the next-year forecast. If we want to use both, we need a suitable restriction.

We want to find a linear function of the expected squared forecast errors of both fixed-event forecasts, i.e. of G_τ and $G_{\tau+1}$, where τ denotes the current year, to approximate the expected squared forecast error of the fixed-horizon forecast of interest. To this end, we need to find the coefficients $b_{1,n}$ and $b_{2,n}$ in the equation

$$E\left[\hat{\varepsilon}^2_{Y_{\lfloor\tilde{s}\rfloor+h}|\tilde{s}}\right] \approx b_{1,n} \times E\left[\hat{\varepsilon}^2_{G_\tau|\tilde{s}}\right] + b_{2,n} \times E\left[\hat{\varepsilon}^2_{G_{\tau+1}|\tilde{s}}\right]$$

where $Y_{\lfloor\tilde{s}\rfloor+h}$ is the fixed-horizon forecast of interest and $n = 1, 2, \ldots, 12$ denotes the month when the forecast is made. The forecast horizon h can have a monthly or quarterly frequency in what follows, and $\lfloor\tilde{s}\rfloor$ simply equals $\tilde{s}$ in the case of a monthly frequency and otherwise equals the quarter containing $\tilde{s}$. The restriction required to use both fixed-event uncertainties could, for example, be chosen as $b_{1,n} + b_{2,n} = 1$. While this specification works reasonably well empirically, it could, in principle, lead to negative values of $E\left[\hat{\varepsilon}^2_{Y_{\lfloor\tilde{s}\rfloor+h}|\tilde{s}}\right]$.[16] Therefore, we employ the restriction

$$b_{1,n} = \lambda\, b_{2,n} \tag{17}$$

with $\lambda \geq 0$, which guarantees non-negative values of both coefficients and contains an additional parameter, which can help to deal with seasonality. A value of $\lambda < 1$ implies that changes in next-year forecast uncertainty have a stronger impact on the fixed-horizon

[16] The restriction $b_{1,n} + b_{2,n} = 1$ corresponds to the restriction used in Ganics et al. (2020), who combine density forecasts. Their coefficients are restricted to be positive. However, for the expected squared forecast error, we find many empirically relevant situations where one of the coefficients is negative. Note that this implies that the density combination proposed by Ganics et al. (2020) could result in densities with misspecified variance.

forecast uncertainty than changes in current-year forecast uncertainty.[17] We could have formulated the restriction as $b_{2,n} = \lambda\, b_{1,n}$, but empirically, $b_{1,n}$ will often tend to be closer to zero than $b_{2,n}$. Therefore, we prefer the specification (17), which allows for the possibility that $b_{1,n} = 0$.

We assume a simple monthly AR(1)–process for g_s in (13) given by

$$g_s = \rho\, g_{s-1} + \varepsilon_s. \tag{18}$$

If this specification was correct, we could set λ to an arbitrary non-negative value. For instance, suppose that $\rho = 0$ and $\sigma^2_{\tilde{s}} = 1$, that forecasts are made in April (i.e. $n = 4$) and that realizations until March are known, such that $\tilde{s}$ corresponds to March of the current year. Then the variance of the forecast error of the current-year growth rate $E\left[\hat{\varepsilon}^2_{G_\tau|\tilde{s}}\right]$ equals about 2, the variance of the forecast error of the next-year growth rate $E\left[\hat{\varepsilon}^2_{G_{\tau+1}|\tilde{s}}\right]$ equals about 8, and the variance of the forecast error of the year-on-year growth rate in March of next year $E\left[\hat{\varepsilon}^2_{Y_{\tilde{s}+11}|\tilde{s}}\right]$ equals 12.[18] Therefore, with $\lambda = 0$, we would obtain $b_{1,4} = 0$ and $b_{2,4} = 1.5$. With $\lambda = 1$, we would obtain $b_{1,4} = 1.2$ and $b_{2,4} = 1.2$. Both sets of parameter values would yield the same fixed-horizon uncertainty $E\left[\hat{\varepsilon}^2_{Y_{\tilde{s}+11}|\tilde{s}}\right] = 12$.

However, the true forecast dynamics can only be approximated by a simple process like (18), implying that different values of λ will lead to different results. Therefore, we search over a grid of values for ρ and λ, determine the associated values of $b_{1,n}$ and $b_{2,n}$ and calculate the resulting fixed-horizon uncertainty. Note that $b_{2,n}$ is simply given by

$$b_{2,n} = \frac{E\left[\hat{\varepsilon}^2_{Y_{\lfloor\tilde{s}\rfloor+h}|\tilde{s}}\right]}{\lambda\, E\left[\hat{\varepsilon}^2_{G_\tau|\tilde{s}}\right] + E\left[\hat{\varepsilon}^2_{G_{\tau+1}|\tilde{s}}\right]}.$$

The quantities $E\left[\hat{\varepsilon}^2_{Y_{\lfloor\tilde{s}\rfloor+h}|\tilde{s}}\right]$, $E\left[\hat{\varepsilon}^2_{G_\tau|\tilde{s}}\right]$ and $E\left[\hat{\varepsilon}^2_{G_{\tau+1}|\tilde{s}}\right]$ are determined based on (15), where only the vector of weights $\boldsymbol{\omega}$ differs between these quantities. σ^2_s is set to 1 without

[17] This result holds for absolute changes, but also for relative changes (in percent), since next-year forecast uncertainty is virtually always larger than current-year forecast uncertainty, which is in line with the assumption of a random-walk behavior of the variance.

[18] The behavior of current-year, next-year, and fixed-horizon uncertainty for different values of ρ and the month of the forecast n is depicted in Figure 4.15 in Appendix 4.8.2.

loss of generality, because we only want to determine $b_{1,n}$ and $b_{2,n}$. These are not affected by the level of σ_s^2 due to the random-walk assumption for the variance.

For each pair of coefficients $b_{1,n}$ and $b_{2,n}$, we then employ the empirical uncertainties using the formula

$$\left(\hat{\sigma}^z_{H|t,n}\right)^2 = b_{1,n}\left(\tilde{\sigma}^z_{t|t,n}\right)^2 + b_{2,n}\left(\tilde{\sigma}^z_{t+1|t,n}\right)^2 \tag{19}$$

to arrive at an estimate of the fixed-horizon uncertainty $\hat{\sigma}^z_{H|t,n}$ with $z \in \{\pi, y\}$, where $\tilde{\sigma}^z_{t|t,n}$ and $\tilde{\sigma}^z_{t+1|t,n}$ are the break-adjusted current year and next year uncertainty from (10) and (12), and π and y denote GDP deflator growth and real GDP growth, respectively. H denotes the target period, which can be a month or quarter. In the monthly case, it is given by $H = (t, n) + h$, where (t, n) denotes the month n of year t when the forecast is made, and h is the forecast horizon in months. In the quarterly case, $H = (t, i) + h$, where (t, i) denotes the quarter i of year t containing the month n when the forecast is made, and h is the forecast horizon in quarters.[19]

Having obtained a series $\hat{\sigma}^z_{H|t,n}$ for each pair of ρ and λ, we calculate the statistic of the Friedman test (see Friedman, 1937) as

$$Q = 0.6\, T \sum_{i=1}^{4} \left(\bar{r}_{\cdot i} - 2.5\right)^2$$

with

$$\bar{r}_{\cdot i} = \frac{1}{T}\sum_{t=1}^{T} r_{ti}$$

where r_{ti} denotes the rank of $\hat{\sigma}^z_{H|t,n}$ within year t, such that r_{ti} takes a value between 1 and 4. T denotes the sample size, and i corresponds to the quarter containing the month n when the forecast is made. The test statistic Q approximately follows a chi-squared distribution with 3 degrees of freedom. We choose the Friedman test because Ollech and Webel (2023), in their comprehensive analysis of several seasonality tests, find it to be one of the most accurate tests. Moreover, it is hardly affected by persistent changes in

[19] (t, n) corresponds to $\tilde{s}$ and t corresponds to τ. We change the notation here, because the approach described above using $\tilde{s}$ and τ is very general, but from now on, we focus on the empirical uncertainty of the US SPF and the ECB SPF.

the level of the time series under study, because it is based on ranks.[20] Such a change will be observed in the ECB SPF's uncertainty.

Having obtained a series $\hat{\sigma}^{z}_{H|t,n}$ for each pair of ρ and λ considered, we choose the pair which yields the lowest Q statistic, labeled Q^*. This value provides the least evidence against the null hypothesis of non-seasonality. Labeling the respective parameter values by ρ^* and λ^*, and the resulting coefficients by $b^*_{1,n}$ and $b^*_{2,n}$, we can compute the approximated fixed-horizon uncertainty series as

$$\left(\hat{\sigma}^{z,*}_{H|t,n}\right)^2 = b^*_{1,n}\left(\tilde{\sigma}^{z}_{t|t,n}\right)^2 + b^*_{2,n}\left(\tilde{\sigma}^{z}_{t+1|t,n}\right)^2.$$

When calculating Q, we ignore years with missing values. When trying to determine ρ^* and λ^*, Q^* might attain a minimum for a set of pairs of ρ and λ. In such a case, we choose the midpoints of the intervals for each parameter as the optimal values ρ^* and λ^*.[21]

4.4.2 An Application to ECB SPF Uncertainty Forecasts

The ECB SPF has been conducted since 1999q1. It asks respondents for forecasts of, among other variables, HICP inflation and real GDP growth in the euro area in the first month of each quarter. Histograms are elicited for both, fixed-event and fixed-horizon forecasts. For inflation and growth, the survey asks for current-year and next-year histogram forecasts like the US SPF. In addition, respondents provide probabilities for their 11-month ahead inflation forecast and their 2-quarter-ahead growth forecasts. These apparently distinct horizons were chosen by the ECB because in both cases the target variable has a distance of one year to the last available value of that variable when the survey is made. For instance, during the January survey, the last available inflation rate is from December of the previous year, and the last available growth rate is from the third quarter of the previous year. Consequently, in January respondents are asked for

[20] The Friedman test can, of course, be applied to any time series where such a structural change takes place around the turn of the year. The test also remains valid if such structural changes occur at random dates. However, if the number of changes at random dates is small, the test might suffer from small sample issues.

[21] The resulting series $\hat{\sigma}^{z,*}_{H|t,n}$ turn out to be very robust to variations of ρ^* and λ^* within the set of Q-minimizing values.

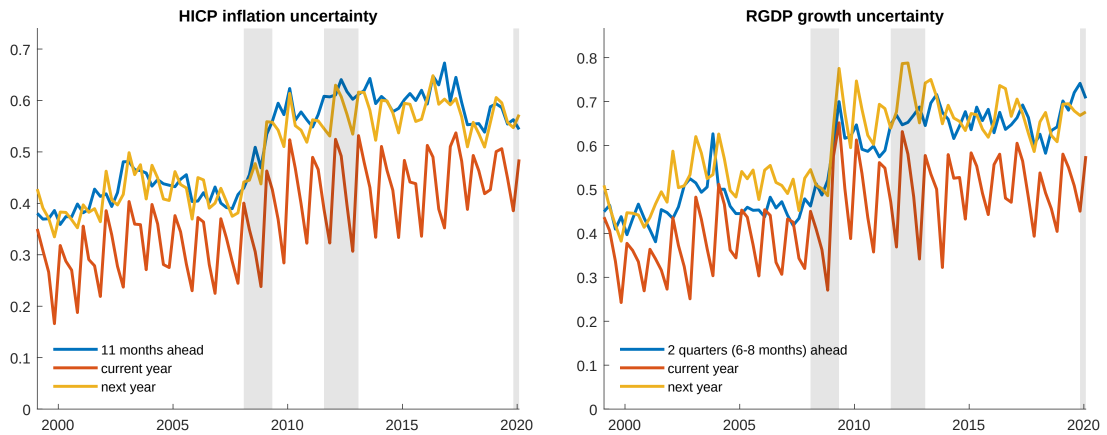


Figure 4.7: Fixed-event and fixed-horizon uncertainty forecasts from the ECB SPF for HICP inflation (left panel) and real GDP growth (right panel). Shaded areas indicate recessions as dated by the CEPR.

their inflation forecast for December of the current year, and for their growth forecast for the third quarter of the current year. Both target variables are year-on-year rates.

From 1999q1 to 2020q1, the widths of the histograms did not change, being equal to 0.5 pp for both variables.[22] The bins became wider for growth during the pandemic. Since we only want to evaluate our seasonal adjustment approach, we restrict our application to the sample ending in 2020q1. The fixed-event and fixed-horizon uncertainty forecasts are shown in Figure 4.7. Apparently, uncertainty increased for all forecasts during the financial crisis and, somewhat surprisingly, did not return to its previous levels until the end of our sample in 2020q1. The co-movement of fixed-horizon and next-year uncertainty is remarkable. Therefore, in Section 4.6, we will apply a standard seasonal adjustment procedure to the next-year uncertainty series (and to the current-year uncertainty series, for the sake of completeness) in the US SPF as an alternative to our approach for seasonal adjustment. However, in this section, we will only focus on our approach.

We search for ρ^* over the grid $\{0, 0.01, 0.02, \ldots, 0.99\}$ and for λ^* over the grid $\{0, 0.01, 0.02, \ldots, 2\}$. We also consider negative values of ρ, but it turns out that Q hardly varies for values of $\rho < 0$. Results for both specifications are shown in Table 4.6.

[22]The number of bins changed over time. The evolution of the bin definitions over time is shown in Figure 4.21 in Appendix 4.8.2.

Table 4.6: Approximation of ECB SPF fixed-horizon uncertainty

	HICP inflation		real GDP growth	
	$\rho \geq 0$	ρ unr.	$\rho \geq 0$	ρ unr.
Q^*	2.20	1.29	2.31	2.09
ρ^*	0.06	-0.42	0.00	-0.24
λ^*	0.26	0.29	0.20	0.20
$b^*_{1,1}$	0.34	0.39	0.17	0.17
$b^*_{2,1}$	1.29	1.35	0.85	0.87
$b^*_{1,4}$	0.36	0.43	0.18	0.18
$b^*_{2,4}$	1.40	1.48	0.90	0.92
$b^*_{1,7}$	0.40	0.47	0.19	0.20
$b^*_{2,7}$	1.53	1.61	0.97	0.99
$b^*_{1,10}$	0.47	0.54	0.23	0.23
$b^*_{2,10}$	1.80	1.88	1.14	1.16
RMSE	0.141	0.161	0.065	0.070
corr.	0.979	0.979	0.919	0.918

Note: ρ unr. refers to searching for the AR-coefficent ρ^* over the set $\{-0.99, -0.98, \ldots, 0.99\}$, whereas $\rho \geq 0$ refers to the set $\{0, 0.01, \ldots, 0.99\}$. λ^* comes from the set $\{0.01, 0.02, \ldots, 2\}$. 'RMSE' denotes the root mean squared error, where the error is measured by the difference between the approximated and the true fixed-horizon uncertainty. 'corr.' denotes the correlation between the approximated and the true fixed-horizon uncertainty. HICP inflation uncertainty refers to 11 months ahead, real GDP growth uncertainty to 2 quarters ahead. The 5% critical value for Q equals 7.81.

While marginally smaller values of Q^* can be obtained with $\rho < 0$, the resulting fixed-horizon uncertainty with $\rho \geq 0$ tends to be closer to the true series. The approximation error, measured by the root of the mean-squared error (RMSE), becomes marginally smaller if $\rho \geq 0$ is imposed. The correlation between approximated and true uncertainty remains basically unaffected by imposing $\rho \geq 0$. Q approximately follows a chi-squared distribution with 3 degrees of freedom, which has a 5% critical value of 7.81, such that all approximated series can be regarded as non-seasonal. These results suggest that considering $\rho < 0$ is not helpful, or might even lead to overfitting.[23]

In any case, it turns out that the approach prefers relatively small values of λ, which implies a larger role for next-year than for current-year uncertainty forecasts in the approximation. Moreover, the values of ρ chosen mostly do not differ much from zero, which implies that the monthly processes of inflation and growth are not persistent.[24] The coefficients for the approximation equation (19) are not too sensitive concerning the use of the restriction $\rho \geq 0$. For instance, with $\rho \geq 0$ and HICP forecasts made in January, we would multiply the current-year uncertainty by $b^*_{1,1} = 0.34$ and the next-year uncertainty by $b^*_{2,1} = 1.29$ to arrive at the 11-month-ahead uncertainty, where uncertainty is measured by the variance. With $\rho < 0$, the respective coefficients would attain similar values, being equal to $b^*_{1,1} = 0.39$ and $b^*_{2,1} = 1.35$. The coefficients differ noticeably between inflation and growth, which is mainly due to the different forecast horizons.

In Figure 4.8, we show how Q depends on ρ and λ. For each graph, we set one parameter to its optimal value and let the other parameter vary. Obviously, λ affects Q noticeably over the entire set of potential values, and with values larger than about 0.6, there would be statistical evidence against the non-seasonality of the approximated inflation and growth uncertainty. However, too small values would also be problematic, at least in the case of inflation. In contrast to that, values of ρ smaller than about 0.3 always lead to non-seasonal behaviour, and there are no smaller values of Q than with $\rho = 0$. In Figure 4.16 in Appendix 4.8.2, we show the same graphs for the case without the restriction $\rho \geq 0$ imposed. They illustrate that Q hardly varies with ρ if $\rho \leq 0$.

[23] We also tried out more complicated processes than AR(1)-processes, namely AR(2)-processes, ARMA(1,1)-processes and ARMA(2,1)-processes. Since they have more parameters than an AR(1)-process, (marginally) lower values of Q^* can be obtained, but the RMSE becomes noticeably larger, which also points to overfitting.

[24] Note that a similar persistence result is obtained in Knüppel and Vladu (2016).

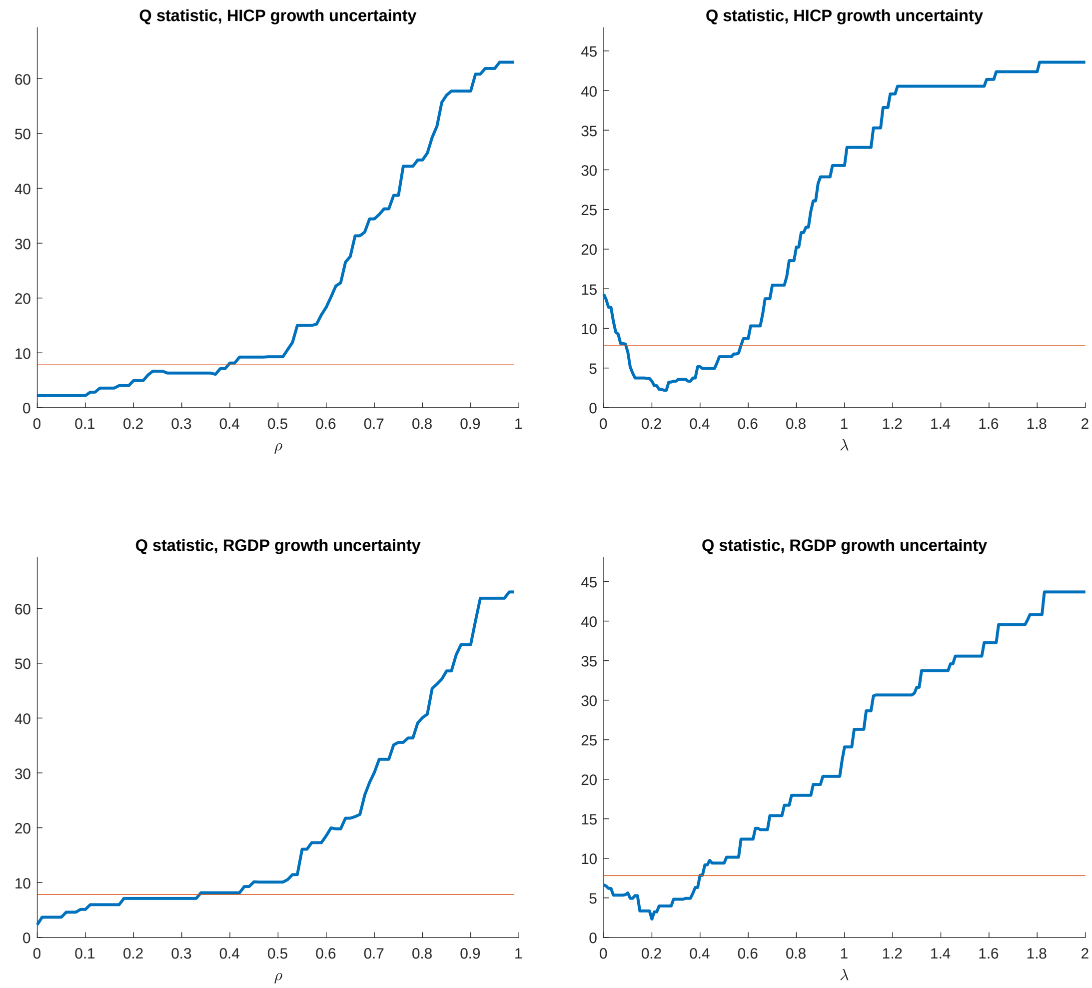


Figure 4.8: Q depending on ρ (left panels) and λ (right panels) for ECB SPF HICP inflation (upper panels) and real GDP growth (lower panels). The Q-values for ρ are obtained by setting $\lambda = \lambda^*$, and the Q-values for λ by setting $\rho = \rho^*$. The red line indicates the 5% critical value of the chi-squared distribution with 3 degrees of freedom, which Q approximately follows.

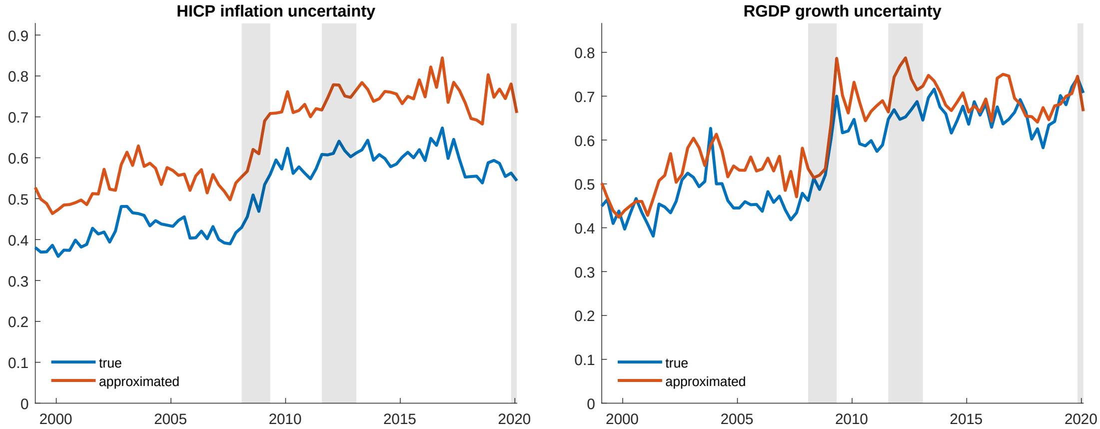


Figure 4.9: True and approximated fixed-horizon uncertainty forecasts from the ECB SPF for HICP inflation (left panel) and real GDP growth (right panel). The approximated uncertainty is derived from the current-year and next-year uncertainty series. Shaded areas indicate recessions as dated by the CEPR.

The approximated fixed-horizon uncertainty series, i.e. the series derived from the fixed-event uncertainty series are displayed in Figure 4.9. Note that the derivation of the fixed-horizon uncertainty series does not make use of any of the information contained in the true fixed-horizon series. Obviously, the approximated inflation uncertainty has a moderate upward bias, while it is strongly correlated with the true uncertainty. The approximated growth uncertainty has a minor upward bias, and its correlation with true uncertainty is not as strong as for inflation, but still pronounced. The figures for $\rho < 0$ can be found in Appendix 4.8.2.

The question might arise why our approach tends to prefer small values of λ. One explanation could be that our approach can explain the next-year uncertainty better than the current-year uncertainty. This, in turn, could be due to horizon-specific biases in forecast uncertainty as documented by Clements (2014), which could be more problematic for the term structure of current-year uncertainty. Therefore, we also calculated bias-adjusted current-year, next-year and fixed-horizon uncertainty. To this end, the original (aggregate) uncertainty series are simply rescaled by a factor, such that for each forecast horizon the rescaled aggregate uncertainty equals the root of the squared forecast errors averaged over time and respondents. However, this procedure leads to values of λ close to zero as well, and the accuracy of the approximation of the true (rescaled) fixed-horizon uncertainty is similar to the case without rescaling. Therefore, we conclude that accounting for horizon-specific biases is neither the reason for low values of λ, nor is it helpful for approximating fixed-horizon uncertainty.

4.5 Seasonal Adjustment for the US SPF

Based on the experiences with the ECB SPF, we derive the fixed-horizon uncertainty forecasts for inflation and growth in the US SPF. We choose a forecast horizon of 3 quarters for both variables.[25] This basically corresponds to the horizon used in the ECB SPF for inflation. For growth, this is one quarter more than in the ECB SPF, but the 3-quarter horizon actually coincides with the ECB SPF's idea to forecast the 4 quarters following the latest available value, since the US SPF takes place one month later, and respondents thus know the previous quarter's release.

Again, we assume an AR(1)-process for the monthly processes, and we impose the restriction of the AR-coefficient being positive. The results obtained are displayed in Table 4.7. Like in the case of the ECB SPF, the values Q^* indicate that the resulting fixed-horizon uncertainty series do not have seasonal patterns. Thus, although the US sample is about twice as large as the ECB sample, it does not appear necessary to consider time-varying parameters of the underlying monthly process (i.e. time variation of ρ) or of the combination approach (i.e. time variation of λ).

The values of ρ^* close to 0 again indicate a very low persistence of the underlying monthly processes. The values of λ^* imply that the next-year uncertainty has a stronger effect on the 3-quarter-ahead uncertainty than the current-year uncertainty. While the fixed-horizon inflation uncertainty approximation almost entirely relies on the next-year uncertainty, the current-year uncertainty still plays an important role for fixed-horizon growth uncertainty. For instance, to arrive at the 3-quarter-ahead uncertainty based

[25] To be more precise, since we assume monthly processes for inflation and growth, we choose a forecast horizon of 11 months and assume that the value in the previous month is known when the survey is made. Note that the choice of a certain fixed horizon will affect the level of the approximated fixed-horizon uncertainty, but not its dynamics. This follows from our assumption of a random-walk behavior for the variance, i.e. from (16). However, the similarity of the 3-quarter-horizon to the fixed horizons used in the ECB SPF, where our seasonal adjustment worked well, makes us prefer this choice over other horizons.

Table 4.7: Approximation of US SPF fixed-horizon uncertainty

	GDP deflator growth	real GDP growth
Q^*	3.44	0.47
ρ^*	0.00	0.19
λ^*	0.14	0.50
$b^*_{1,2}$	0.16	0.48
$b^*_{2,2}$	1.12	0.96
$b^*_{1,5}$	0.16	0.54
$b^*_{2,5}$	1.18	1.08
$b^*_{1,8}$	0.18	0.62
$b^*_{2,8}$	1.28	1.24
$b^*_{1,11}$	0.22	0.79
$b^*_{2,11}$	1.58	1.57

Note: For the AR-coefficient $\rho \geq 0$ is imposed. Inflation and growth uncertainty refer to 3-quarter-ahead forecasts. The 5% critical value for Q equals 7.81.

on the November survey, one needs to multiply the next-year uncertainty by about 1.6 for both, inflation and growth. For inflation, however, one only adds about 0.1 times the current-year uncertainty, while for growth, one adds 0.8 times the current-year uncertainty. Figure 4.10 shows that values of ρ between 0 and about 0.3 lead to a non-seasonal fixed-horizon uncertainty approximation for both variables conditional on employing λ^*, respectively. Concerning λ, values lower than about 0.5 are required for GDP deflator growth, while values between about 0.2 and 0.9 are needed for real GDP growth conditional on employing ρ^*, respectively.

The resulting 3-quarter-ahead inflation and growth uncertainty series are shown in Figure 4.11. Obviously, growth uncertainty peaked at the beginning of the pandemic, while inflation uncertainty did in 2021. During all other recessions, growth uncertainty

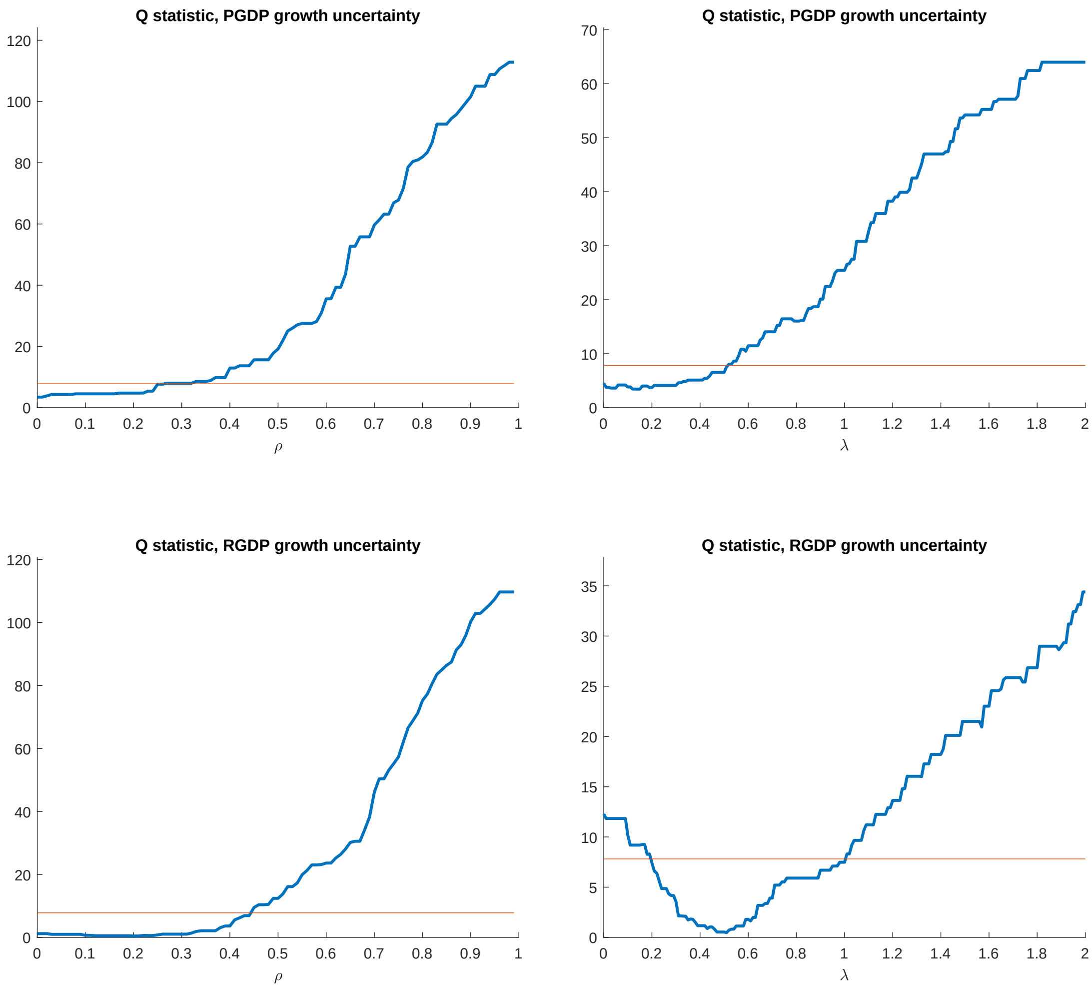


Figure 4.10: Q depending on ρ (left panels) and λ (right panels) for US SPF GDP deflator growth (upper panels) and real GDP growth (lower panels). The Q-values for ρ are obtained by setting $\lambda = \lambda^*$, and the Q-values for λ by setting $\rho = \rho^*$. The red line indicates the 5% critical value of the chi-squared distribution with 3 degrees of freedom, which Q approximately follows.

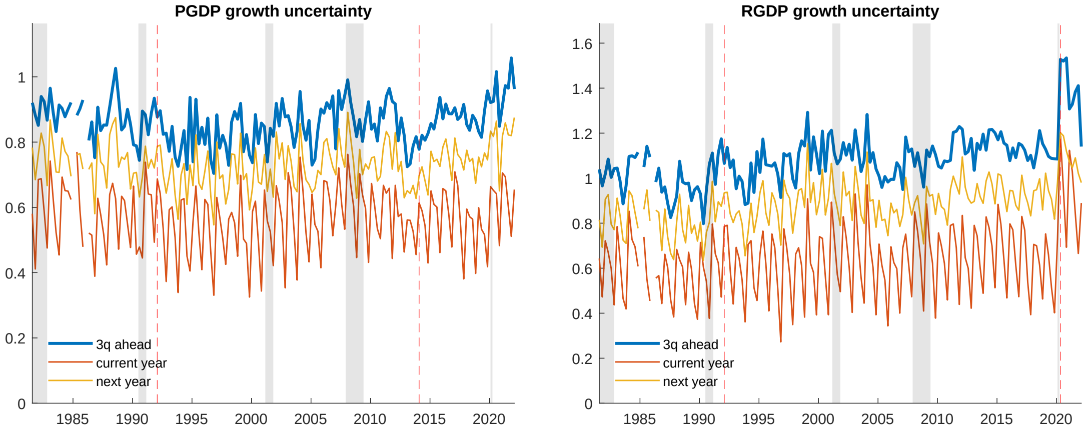


Figure 4.11: Break- and seasonally-adjusted 3-quarter-ahead GDP deflator growth uncertainty (left panel) and real GDP growth uncertainty (right panel). The fixed-event uncertainty series are the break-adjusted series from Figure 4.6.

does not seem to be particularly elevated. Moreover, there is no evident decline in uncertainty in the course of the 1980s. We will discuss this and other properties below.

4.6 Comparison to Related Uncertainty Measures

The evolution of the US SPF's ex ante uncertainty displayed in Figure 4.11 raises the question about its relation to other uncertainty measures like those of Baker et al. (2016), Jurado et al. (2015) or Binder (2017), where the latter focuses on inflation only. Moreover, it seems especially interesting to compare the US SPF's ex ante uncertainty to other uncertainty measures derived from the US SPF, such as the ex post measure by Rossi and Sekhposyan (2015) or the disagreement measure provided by the US SPF, which is based on the respondents' point forecasts.[26]

To make the monthly indices comparable to ours, we use only values from the months when the US SPF was conducted. We employ the Economic Policy Uncertainty (EPU) from Baker et al. (2016). Jurado et al. (2015) offer three types of uncertainty indices (real, macro, financial) with two subtypes each (total and economic, where the difference is due

[26] We focus on the difference between the 75th percentile and the 25th percentile of the quarter-on-quarter growth forecasts. Other dispersion measures provided by the US SPF use the levels of the forecasts.

to the effect of COVID-19 on uncertainty) and three forecast horizons each. We simply use the average across horizons and subtypes, because all six corresponding series are highly correlated for each uncertainty type considered. We focus on Macro Uncertainty only, and we label it JLN. We employ the short-horizon inflation uncertainty data by Binder (2017), which corresponds to a forecast horizon of 1 year, denoting it by B. From Rossi and Sekhposyan's (2015) uncertainty series, we choose the 4-quarter-ahead uncertainty based on revised data because it turns out to have the highest average correlation with EPU and JLN in each of the samples considered below. We label this measure RS. The disagreement measure provided by the US SPF contains the interquartile ranges of point forecasts for the current quarter (the nowcast) and the following 4 quarters. We simply take the average across these horizons, because this average turns out to have a higher average correlation with EPU and JLN than each horizon-specific measure in each of the samples considered below. We consider the disagreement for GDP deflator growth and real GDP growth, and label these uncertainty measures DIS^{π} and DIS^{y}, respectively. Note that this definition of disagreement differs from the definition briefly used in Section 4.1.

In addition to these established uncertainty measures, we also employ a different seasonal adjustment procedure than proposed in Section 4.4 to obtain four additional uncertainty series. We simply apply the Tramo/Seats procedure of Gómez and Maravall (1998) to our break-adjusted current- and next-year uncertainty of inflation and growth from the US SPF.[27] One reason to include these series is the similarity of the fixed-horizon and the next-year uncertainty in the ECB SPF shown in Figure 4.7. The symbols, abbreviations, and expressions to describe the different uncertainty measures are collected in Table 4.8.

Figures 4.17 and 4.18 in Appendix 4.8.2 show the standardized uncertainty measures, where the standardization of an uncertainty measure Z simply means subtracting its mean μ_Z and then dividing by its standard deviation σ_Z. Obviously, the pandemic led to high levels of uncertainty. In the cases of EPU and DIS^{y}, the peak in uncertainty is extreme, attaining almost 7 (EPU) and 10 (DIS^{y}) standard deviations, respectively.

[27] We use the default implementation of Tramo/Seats for quarterly time series in EViews 13.

Table 4.8: Uncertainty measures

$\hat{\sigma}^{\pi,*}$	FH inflation uncertainty	approximated fixed-horizon GDP deflator growth uncertainty
$\hat{\sigma}^{y,*}$	FH growth uncertainty	approximated fixed-horizon real GDP growth uncertainty
$\tilde{\sigma}_t^{\pi,TS}$	CY inflation uncertainty (TS)	current-year GDP deflator growth uncertainty, seasonally adjusted with Tramo/Seats
$\tilde{\sigma}_{t+1}^{\pi,TS}$	NY inflation uncertainty (TS)	next-year GDP deflator growth uncertainty, seasonally adjusted with Tramo/Seats
$\tilde{\sigma}_t^{y,TS}$	CY growth uncertainty (TS)	current-year real GDP growth uncertainty, seasonally adjusted with Tramo/Seats
$\tilde{\sigma}_{t+1}^{y,TS}$	NY growth uncertainty (TS)	next-year real GDP growth uncertainty, seasonally adjusted with Tramo/Seats
DIS^{π}	inflation disagreement	interquartile ranges of point forecasts for GDP deflator growth, averaged over the current and next four quarters
DIS^{y}	growth disagreement	interquartile ranges of point forecasts for real GDP growth, averaged over the current and next four quarters
RS	ex-post uncertainty	four-quarter-ahead ex post uncertainty based on revised data; Rossi and Sekhposyan (2015)
EPU	Economic Policy Uncertainty	Economic Policy Uncertainty index based on newspaper coverage frequency; Baker et al. (2016).
JLN	Macro Uncertainty	Macro Uncertainty following Jurado et al. (2015), averaged across horizons and types (total and real)
B	rounding-based inflation uncertainty	one-year-ahead inflation uncertainty based on the rounding patterns of point forecasts (multiples of five) from the Michigan Survey of Consumers; Binder (2017)

Note: The first 6 series rely on the break adjustments described in Section 4.3.

4.6.1 Correlations

Table 4.9 shows the correlations between all uncertainty measures in three different samples. The correlation between the uncertainty series A_s and B_s is obtained by running the regression

$$\frac{(A_s - \mu_A)\,(B_s - \mu_B)}{\sigma_A\,\sigma_B} = \alpha + u_s, \tag{20}$$

with u_s being the residual and $s = s_{min}, s_{min} + 1, \ldots, s_{max}$. The correlation coefficient is given by α, and the regression allows us to calculate its autocorrelation-consistent standard error. We employ Newey and West (1987) standard errors with truncation lags chosen as proposed by Andrews (1991).

The first of the three rows of correlations always refers to the largest sample from 1981q3 to 2022q1. In this sample, correlations can be expected to be relatively large, because all uncertainty measures were at high levels during the first quarters of the pandemic. The second sample starts in 1981q3 as well but ends in 2019q4 and, thus, before the pandemic started. The last sample starts in 1992q1 and ends in 2019q4. This sample is unaffected by structural breaks due to new bin definitions for real GDP growth in the US SPF. Moreover, it only contains one such break for real GDP deflator growth, which we adjusted based on information from other US SPF variables for inflation. Therefore, even if our adjustment for structural breaks might be problematic for the breaks observed in 1992q1 or 2020q2, the results in the third sample should be informative about the behavior of ex ante uncertainty in the US SPF. Only the well-established uncertainty measures Economic Policy Uncertainty, Macro Uncertainty, and rounding-based inflation uncertainty are not based on the US SPF.[28]

All six ex ante uncertainty measures from the US SPF except for disagreement, namely FH growth and inflation uncertainty as well as CY and NY inflation and growth uncertainty (TS) are significantly, and often strongly correlated with each other, which is not surprising at least for FH growth and inflation uncertainty due to their construction.[29] However, most US SPF uncertainty measures are only weakly and insignificantly correlated with Economic Policy Uncertainty and Macro Uncertainty, especially in the samples ending before the pandemic, while the correlations with rounding-based inflation uncertainty

[28] EPU uses some disagreement measures for policy variables from the US SPF among many other data.
[29] Please refer to Table 4.8 for the abbreviations and terms.

Table 4.9: Correlations of uncertainty measures

	$\hat{\sigma}^{y,*}$	$\tilde{\sigma}_t^{\pi,TS}$	$\tilde{\sigma}_{t+1}^{\pi,TS}$	$\tilde{\sigma}_t^{y,TS}$	$\tilde{\sigma}_{t+1}^{y,TS}$	DIS^{π}	DIS^{y}	RS	EPU	JLN	B
$\hat{\sigma}^{\pi,*}$	0.41***	0.69***	0.99***	0.43**	0.38***	0.29***	0.22*	0.02	0.18*	0.31**	0.22***
$\hat{\sigma}^{\pi,*}$	0.33***	0.69***	0.99***	0.34***	0.29***	0.25***	0.17*	0.00	0.08	0.20**	0.21***
$\hat{\sigma}^{\pi,*}$	0.50***	0.72***	0.99***	0.43***	0.47***	0.10	−0.02	0.04	0.07	0.19**	0.15*
$\hat{\sigma}^{y,*}$		0.39***	0.41***	0.84**	0.97***	−0.11	0.27	−0.04	0.37	0.25	0.07
$\hat{\sigma}^{y,*}$		0.34***	0.33***	0.73***	0.97***	−0.25***	−0.32**	−0.16	0.00	−0.12	−0.04
$\hat{\sigma}^{y,*}$		0.41***	0.51***	0.75***	0.97***	−0.07	0.02	−0.03	0.05	−0.05	−0.03
$\tilde{\sigma}_t^{\pi,TS}$			0.68***	0.51***	0.32***	0.34***	0.19*	−0.11	0.23**	0.29***	0.36***
$\tilde{\sigma}_t^{\pi,TS}$			0.67***	0.55***	0.25**	0.34***	0.17	−0.13	0.17*	0.25**	0.35***
$\tilde{\sigma}_t^{\pi,TS}$			0.72***	0.57***	0.33***	0.25**	0.05	−0.10	0.14	0.31*	0.34***
$\tilde{\sigma}_{t+1}^{\pi,TS}$				0.43**	0.38***	0.29***	0.23*	0.03	0.18	0.31**	0.21***
$\tilde{\sigma}_{t+1}^{\pi,TS}$				0.33***	0.29***	0.25***	0.17*	0.02	0.07	0.19**	0.20***
$\tilde{\sigma}_{t+1}^{\pi,TS}$				0.44***	0.48***	0.10	−0.02	0.05	0.05	0.18**	0.14*
$\tilde{\sigma}_t^{y,TS}$					0.73**	0.11	0.42	−0.05	0.45	0.42	0.26**
$\tilde{\sigma}_t^{y,TS}$					0.60***	0.03	−0.09	−0.20**	0.09	0.09	0.22**
$\tilde{\sigma}_t^{y,TS}$					0.63***	0.19	0.17	−0.17	0.16	0.16	0.25**
$\tilde{\sigma}_{t+1}^{y,TS}$						−0.17*	0.18	−0.05	0.29	0.17	0.00
$\tilde{\sigma}_{t+1}^{y,TS}$						−0.29***	−0.34***	−0.14	−0.04	−0.15*	−0.09
$\tilde{\sigma}_{t+1}^{y,TS}$						−0.12	−0.01	0.02	0.02	−0.09	−0.09
DIS^{π}							0.49***	0.07	0.38***	0.45***	0.61***
DIS^{π}							0.74***	0.07	0.39***	0.50**	0.61***
DIS^{π}							0.57**	−0.05	0.38**	0.46	0.69***
DIS^{y}								0.23*	0.70	0.65*	0.46***
DIS^{y}								0.21	0.39***	0.56**	0.57**
DIS^{y}								0.23**	0.33*	0.62	0.74**
RS									0.12	0.29**	0.13
RS									−0.01	0.25*	0.11
RS									0.03	0.16	0.01
EPU										0.52	0.47***
EPU										0.23	0.44***
EPU										0.23	0.48***
JLN											0.64***
JLN											0.71**
JLN											0.66*

Note: The first of the three rows of correlations always refers to the sample 1981q3 to 2022q1, the second row to 1981q3 to 2019q4, and the third row to 1992q1 to 2019q4. For the symbols, abbreviations and descriptions of the individual uncertainty measures, see Table 4.8. Dates with missing values are dropped from the respective pairwise samples. $\hat{\sigma}^{z,*}$ have missing values in 1985q1, 1986q1, and 1990q1. EPU starts in 1985q1, RS ends in 2021q2. *,**,*** denote a significant difference from 0 at the 10%, 5%, 1% level, respectively. Significance is assessed based on Newey and West (1987) standard errors with truncation lags chosen as proposed by Andrews (1991). See equation (20) for details.

often tend to be significant. Interestingly, FH growth uncertainty does not have any significant correlations with Economic Policy Uncertainty, Macro Uncertainty, or rounding-based inflation uncertainty. In contrast to that, FH inflation uncertainty is weakly but significantly correlated with Macro Uncertainty and rounding-based inflation uncertainty in all samples.

The ex post uncertainty is weakly and, except for the smallest sample, significantly correlated with Macro Uncertainty. However, it is basically uncorrelated with Economic Policy Uncertainty and rounding-based inflation uncertainty in all samples.[30] In contrast to that, inflation and growth disagreement are moderately or strongly correlated with Economic Policy Uncertainty, Macro Uncertainty, and rounding-based inflation uncertainty in all samples, and most of these correlations are significant.[31] Therefore, inflation and growth disagreement appear to be the only uncertainty measures from the US SPF that convey information similar to established uncertainty measures.

4.6.2 VAR analysis

Many theoretical considerations and empirical findings suggest that uncertainty shocks have a negative impact on economic activity. In addition to the studies mentioned above, there is, for instance, a growing literature that documents that higher uncertainty reduces household spending (see Armantier et al., 2021; Coibion et al., 2022a).[32] Therefore, we are going to investigate if our US SPF-based uncertainty measures lead to the expected result in a standard VAR. We follow Rossi and Sekhposyan (2015) and estimate a quarterly VAR with 5 variables ordered as follows: The (standardized) uncertainty measure, the log of the S&P 500, the Federal Funds rate, the log of employment, and the log of real GDP. The VAR also contains a time trend and is estimated from 1985q1 to 2019q4. Missing values in the ex ante uncertainty measures from the US SPF are filled by linear interpolation. We choose 2 lags in each regression because the BIC is minimized by 2 lags in most regressions and by 1 lag in the remaining cases. We study the effect of a one-standard-deviation uncertainty shock to real GDP over a period of 20 quarters. We

[30] Rossi and Sekhposyan (2015) propose additional measures capturing upside and downside uncertainty, which might have larger correlations. However, these measures are special in the sense that they often remain at their minimum of 0.5 over longer periods.

[31] The correlation between growth disagreement and Economic Policy Uncertainty in the largest sample attains a large value of 0.7 but is statistically insignificant because both series have an outlier in 2020q2. Removing this outlier would lead to a correlation of 0.25 and a p-value of 0.01.

[32] Yet, the growth option theories mentioned in Ludvigson et al. (2021) consider the possibility that certain types of uncertainty might have positive effects on economic activity.

focus on uncertainty measures related to economic activity but also consider FH inflation uncertainty.

Figure 4.12 shows the effect of shocks to FH inflation and growth uncertainty, to Economic Policy Uncertainty, and to Macro Uncertainty. The fixed-horizon uncertainty shocks tend to have positive effects on real GDP, although the responses are not significantly different from zero if a 5% significance level is used. In contrast to that, shocks to Economic Policy Uncertainty and to Macro Uncertainty have the expected negative effects, which are significant during the first three quarters for Economic Policy Uncertainty and during the entire 20 quarters considered for Macro Uncertainty.

Figure 4.13 shows the effects of shocks to CY and NY growth uncertainty (TS), to growth disagreement, and to ex post uncertainty. While shocks to CY growth uncertainty (TS) and ex post uncertainty only have very small, insignificant effects on real GDP, shocks to NY growth uncertainty (TS) increase real GDP noticeably, and these increases are often significant. Only shocks to growth disagreement have a negative effect on real GDP for all 20 quarters. While this effect is insignificant, it appears to be the most similar to Economic Policy Uncertainty and Macro Uncertainty among all US SPF-based uncertainty measures considered.[33]

In Appendix 4.8.2 in Figures 4.19 and 4.20, we include the pandemic in the estimation sample, which lasts from 1985q1 to 2022q1. By and large, the responses to uncertainty shocks are shifted upwards. However, the basic results remain unchanged. Economic Policy Uncertainty and Macro Uncertainty yield significantly negative responses of real GDP for at least some quarters, while all US SPF-based uncertainty measures except for growth disagreement give either very small effects or large positive effects. A shock to growth disagreement has a significantly negative effect on impact, but the effect is close to zero for the following periods. The impact result is likely due to the spike in disagreement in 2020q2 and the pronounced drop in real GDP in the same quarter. When reducing the sample to the period 1992q1 to 2019q4, we do not observe negative real GDP responses to FH inflation and growth uncertainty either, while Economic Policy Uncertainty and Macro Uncertainty shocks continue to have this effect.

[33]The GDP effects observed with inflation disagreement, not shown here, tend to be closer to zero than those with growth disagreement.

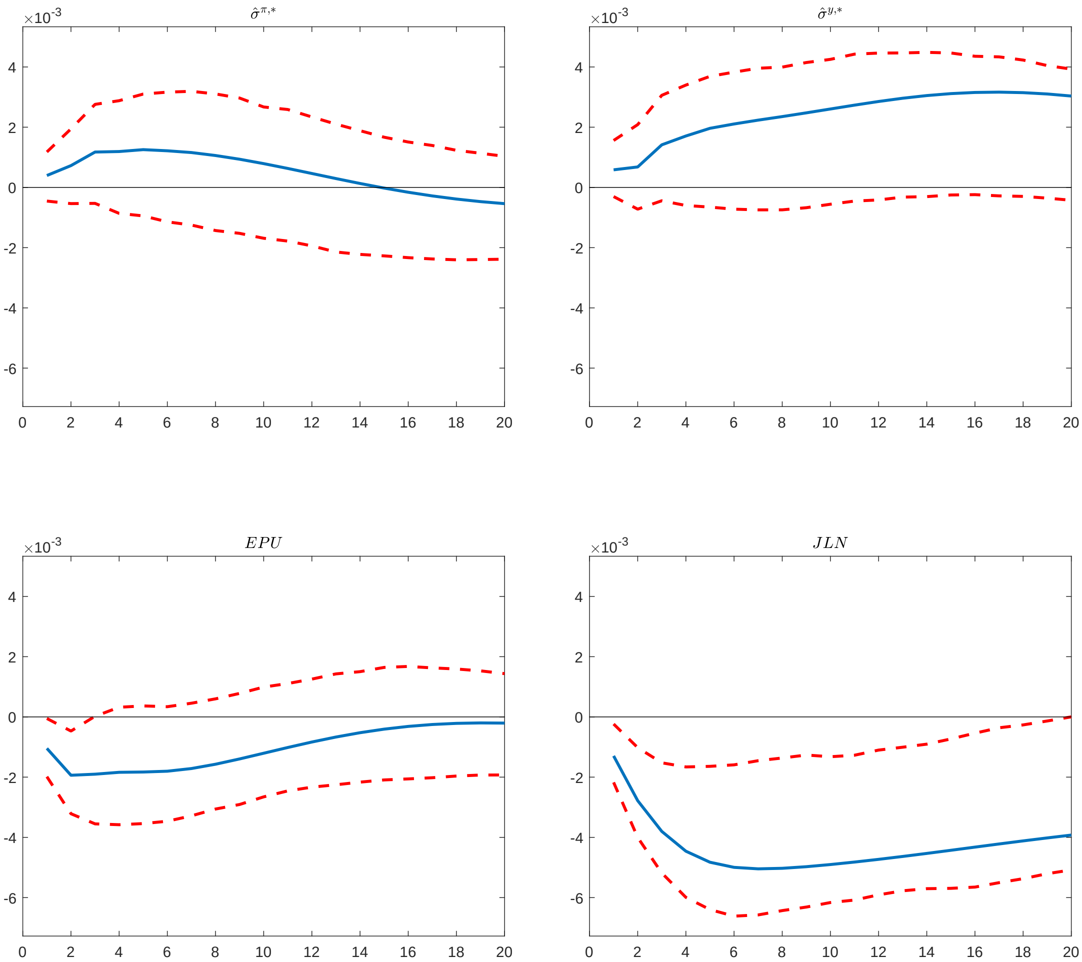


Figure 4.12: Responses of real GDP to a one-standard-deviation uncertainty shock in quarter 1. The red dashed lines are lower and upper 95% confidence bounds for each quarter estimated by bootstrapping. The uncertainty measures are FH inflation uncertainty (top left), FH growth uncertainty (top right), Economic Policy Uncertainty (bottom left), and Macro Uncertainty (bottom right). For the abbreviations and descriptions of the individual uncertainty measures, see Table 4.8. The estimation sample is 1985q1 to 2019q4.

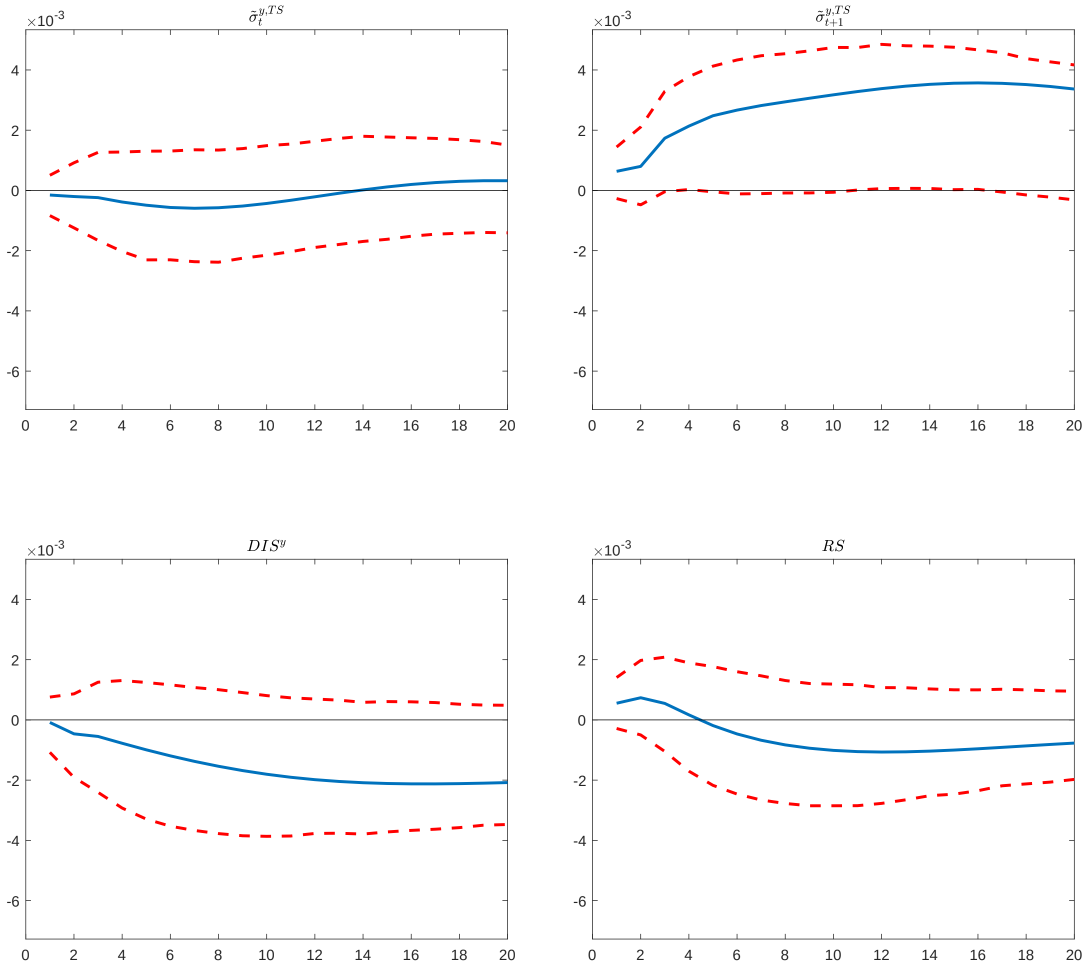


Figure 4.13: Responses of real GDP to a one-standard-deviation uncertainty shock in quarter 1. The red dashed lines are lower and upper 95% confidence bounds for each quarter estimated by bootstrapping. The uncertainty measures are CY (top left) and NY (top right) growth uncertainty, growth disagreement (bottom left), and ex post uncertainty (bottom right). For the abbreviations and descriptions of the individual uncertainty measures, see Table 4.8. The estimation sample is 1985q1 to 2019q4.

4.7 Conclusion

We quantify the ex ante uncertainty for inflation and growth based on the corresponding histogram forecasts of the US SPF. To this end, we first suggest a procedure to adjust the measured ex ante uncertainty for changes in the survey design, namely for changes in the bin widths of the histograms. Then we propose an approach to combine the current- and next-year uncertainty such that the resulting fixed-horizon uncertainty does not have seasonal patterns.

We compare the properties of the resulting fixed-horizon uncertainty to other established uncertainty measures. We find that fixed-horizon uncertainty for growth is hardly correlated with the economic uncertainty measures of Baker et al. (2016) and Jurado et al. (2015) or the inflation uncertainty measure of Binder (2017). The fixed-horizon uncertainty for inflation tends to be weakly, but significantly correlated with the uncertainty measures of Jurado et al. (2015) and Binder (2017). However, the by far strongest correlations of US SPF-based uncertainty measures with those of Baker et al. (2016), Jurado et al. (2015), and Binder (2017) are observed for the disagreement among US SPF respondents concerning their point forecasts.

We use a standard VAR to study the response of economic activity to uncertainty shocks. It turns out that shocks to fixed-horizon uncertainty for growth and inflation do not have the expected effect, i.e. economic activity does not decrease. Among all US SPF-based uncertainty measures, only the VAR with growth disagreement produces responses to uncertainty shocks that are broadly comparable to those obtained in VARs with the uncertainty measure of Baker et al. (2016) or Jurado et al. (2015).

All results concerning the comparison of our fixed-horizon uncertainty to other established uncertainty measures tend to be robust to the choice of different samples or seasonal-adjustment procedures. Excluding the pandemic recession from the sample, in general, leads to less commonality among the uncertainty measures. Most importantly, our results concerning fixed-horizon growth uncertainty also hold if we consider the largest break-free subsample and a standard seasonal-adjustment procedure.

While it is possible that the ex ante uncertainty of individual forecasters indeed behaves very differently from the uncertainty identified by Baker et al. (2016) or Jurado et al. (2015), the large effects of changes in the histogram bin widths on measured ex ante uncertainty suggest that there seem to be important measurement problems associated with the current practice of eliciting probabilistic forecasts. Such measurement problems might also be responsible for the uncommon behavior of ex ante uncertainty. Therefore, we believe that the effects of survey design need to be studied for professional forecasters, just like it is done in the case of consumer surveys (see e.g. Bruine de Bruin et al., 2017). While such studies will inevitably have to confront the difficulty of relatively small samples of professional forecasters, there is, in our view, no reasonable alternative to this approach. Otherwise, the time series derived from histogram forecasts will almost certainly become subject to structural breaks that might render these series useless for many types of analyses. Basically, one needs to understand how changes in histogram definitions affect the probabilities reported by professional forecasters, and one maybe should search for alternatives to histogram forecasts that are robust to large changes in the behavior of the forecast target. For instance, an interesting alternative to histogram forecasts was proposed by Altig et al. (2022). Asking for quantiles would also be a possibility to obtain potentially more robust information about forecast distributions. It would be important to know if such approaches can deliver more reliable results than histogram forecasts when used by professional forecasters.

4.8 Appendix

4.8.1 Details of Quantification Procedure

Cases 1 and 2 deal with forecasters who assign positive probabilities to 1 or 2 bins only, such that a triangular distribution is fitted. Case 3 considers positive probabilities for at least 3 bins, such that a normal distribution can be employed.

Case 1 The forecaster uses a single bin with width w. That is, the interval has the range $(A, A+w]$. Then the support of the triangle contains the entirety of the interval. If the probability mass is allocated in a single open-ended bin, then we assume it has width $2w$, where w is the width of the adjacent closed bin.

Case 2 Assume that the forecaster uses two intervals with width v and w. That is, the intervals have the ranges $(A-v, A]$ and $(A, A+w]$, respectively. Let q_v and q_w be the probabilities assigned to the intervals and B be the base of the triangle.

Case 2.1 $v > w$ i.e. the larger interval is on the left

Case 2.1.1 if $q_w \leq 2\left(\frac{w}{v+w}\right)^2$

One can pin the left endpoint $a = A - v$, then using $F(A) = P(A - v < x \leq A) = q_v = 1 - q_w$ solve for b

$$b = A - \frac{\sqrt{q_w}}{\sqrt{q_w} - \sqrt{2}} v$$

$$B = b - a = v\left(1 - \frac{\sqrt{q_w}}{\sqrt{q_w} - \sqrt{2}}\right)$$

Case 2.1.2 if $2\left(\frac{w}{v+w}\right)^2 \leq q_w \leq \frac{1}{2}$

Then one can set $b = A + w$ and solving for a yields

$$a = (A + w) - \frac{\sqrt{2}}{\sqrt{q_w}} w$$

$$B = b - a = \frac{\sqrt{2}}{\sqrt{q_w}} w$$

Case 2.1.3 if $q_w > \frac{1}{2}$ which means, there is more probability in the right interval and we can proceed as if the intervals are of the same length. We can pin the right endpoint of the support $b = A + w$, and $F(A) = q_v$. Solving for a

$$a = A - \frac{\sqrt{q_v}}{\sqrt{2} - \sqrt{q_v}} w$$

$$B = b - a = \frac{2\sqrt{q_v} - \sqrt{2}}{\sqrt{q_v} - \sqrt{2}} w$$

Case 2.2 $v < w$ *i.e. the larger interval is on the right*

Case 2.2.1 if $q_v \leq 2\left(\frac{v}{v+w}\right)^2$
Then one can let $b = A + w$ and using $F(A) = q_v$

$$a = A + \frac{\sqrt{q_v}}{\sqrt{q_v} - \sqrt{2}} w$$

$$B = b - a = w\left(1 - \frac{\sqrt{q_v}}{\sqrt{q_v} - \sqrt{2}}\right)$$

Case 2.2.2 if $2\left(\frac{v}{v+w}\right)^2 < q_v \leq \frac{1}{2}$
Then one can let $a = A - v$ and solving for b

$$b = A + \frac{\sqrt{2} - \sqrt{q_v}}{\sqrt{q_v}} v$$

$$B = b - a = v\left(\frac{\sqrt{2} - \sqrt{q_v}}{\sqrt{q_v}} - 1\right)$$

Case 2.2.3 if $q_v \geq \frac{1}{2}$
Then we can proceed as if the intervals have the same length. Then we can fix the left endpoint $a = A - v$ as before and solve for b

$$b = A - v\left(\frac{\sqrt{q_w}}{\sqrt{q_w} - \sqrt{2}}\right)$$

$$B = b - a = \left(1 - \frac{\sqrt{q_w}}{\sqrt{q_w} - \sqrt{2}}\right)v$$

Case 2.3 $v = w$, i.e. the two intervals are of equal length

Case 2.3.1 if $q_v \geq \frac{1}{2}$, then we can fix the left endpoint $a = A - v$ and solve for b

$$b = A - v\left(\frac{\sqrt{q_w}}{\sqrt{q_w} - \sqrt{2}}\right)$$

$$B = b - a = \left(1 - \frac{\sqrt{q_w}}{\sqrt{q_w} - \sqrt{2}}\right)v$$

Case 2.3.2 if $q_v < \frac{1}{2}$, then we can fix the right endpoint at $b = A + w$ and solve for a

$$a = A - \frac{\sqrt{q_v}}{\sqrt{2} - \sqrt{q_v}}w$$

$$B = b - a = \frac{2\sqrt{q_v} - \sqrt{2}}{\sqrt{q_v} - \sqrt{2}}w$$

Case 3 The forecasters use three or more bins. In this case, we assume a Normal distribution and minimize

$$\min_{\mu,\sigma} \frac{1}{K} \sum_{k=1}^{K} \left[\Phi\left(x, \mu, \sigma\right) - F(x)\right]^2$$

where K is the number of bins and x contains all upper bounds of the intervals. μ and σ denote the mean and standard deviation, respectively.

4.8.2 Additional Figures

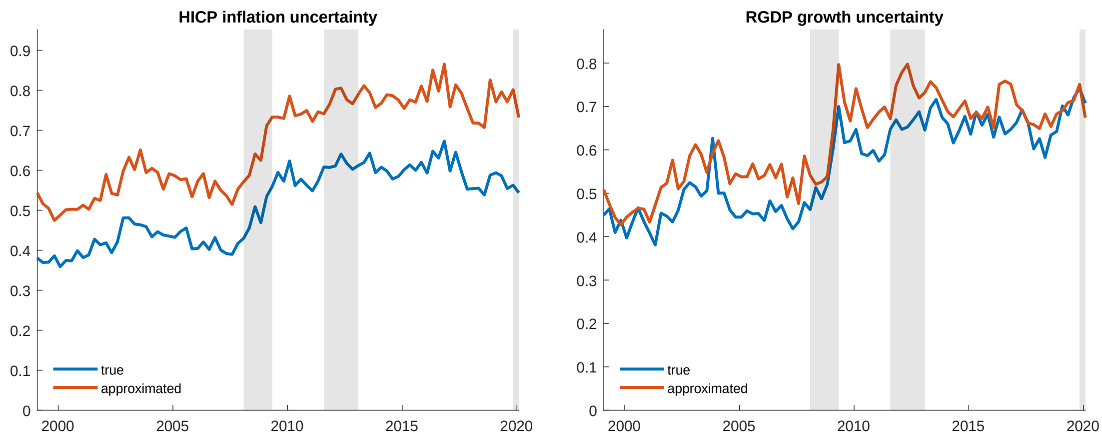


Figure 4.14: True and approximated fixed-horizon uncertainty forecasts from the ECB SPF for HICP inflation (left panel) and real GDP growth (right panel) without the restriction $\rho \geq 0$ imposed. The approximated uncertainty is derived from the current-year and next-year uncertainty series. Shaded areas indicate recessions as dated by the CEPR.

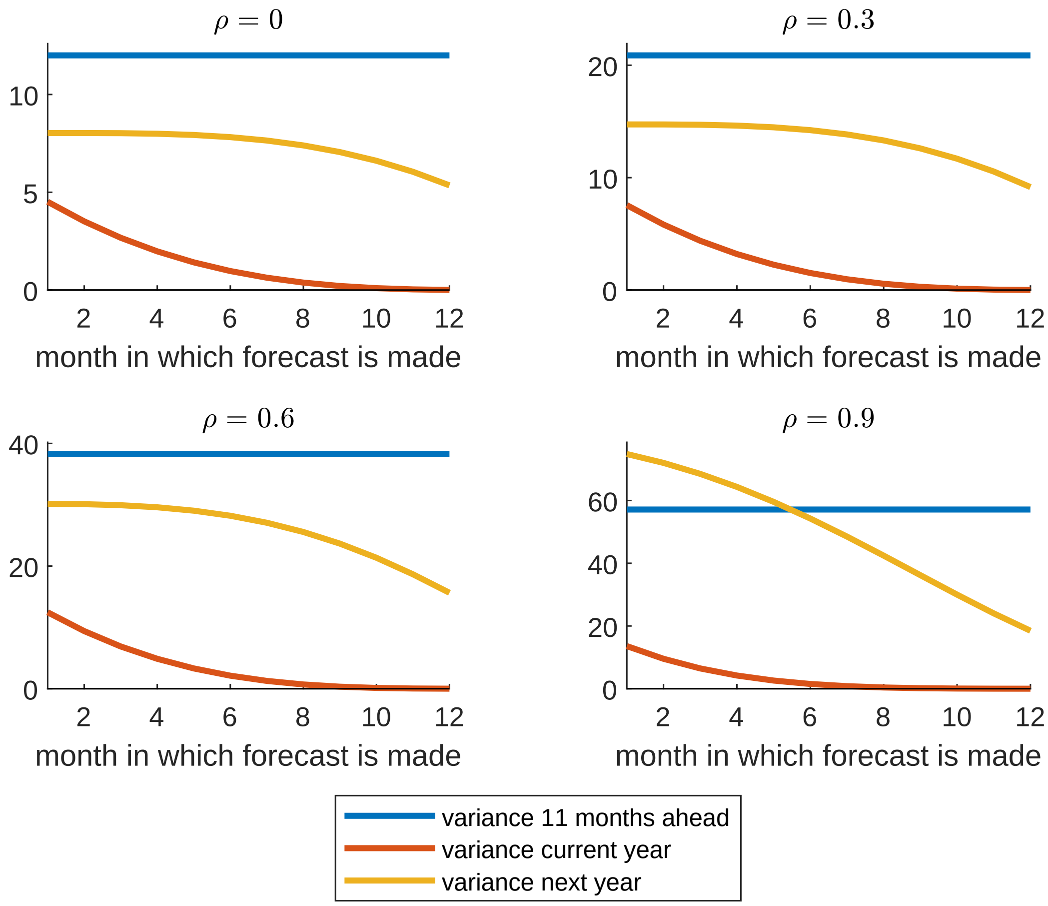


Figure 4.15: Fixed-horizon, current-year, and next-year uncertainty (variance) for different values of ρ based on the assumptions described in Section 4.4. ρ is the AR-coefficient of the AR(1)-process (18), $\sigma^2_{\tilde{s}} = 1$ for each $\tilde{s}$, where $\tilde{s}$ is assumed to be the month before the month when the forecast is made. For instance, when the forecast is made in April, $\tilde{s}$ corresponds to March of the same year, implying that the realizations until March are known in April. The uncertainties correspond to the variance of the forecast error of the 11-month-ahead year-on-year growth rate, the variance of the forecast error of the current-year growth rate, and the variance of the forecast error of the next-year growth rate, respectively.

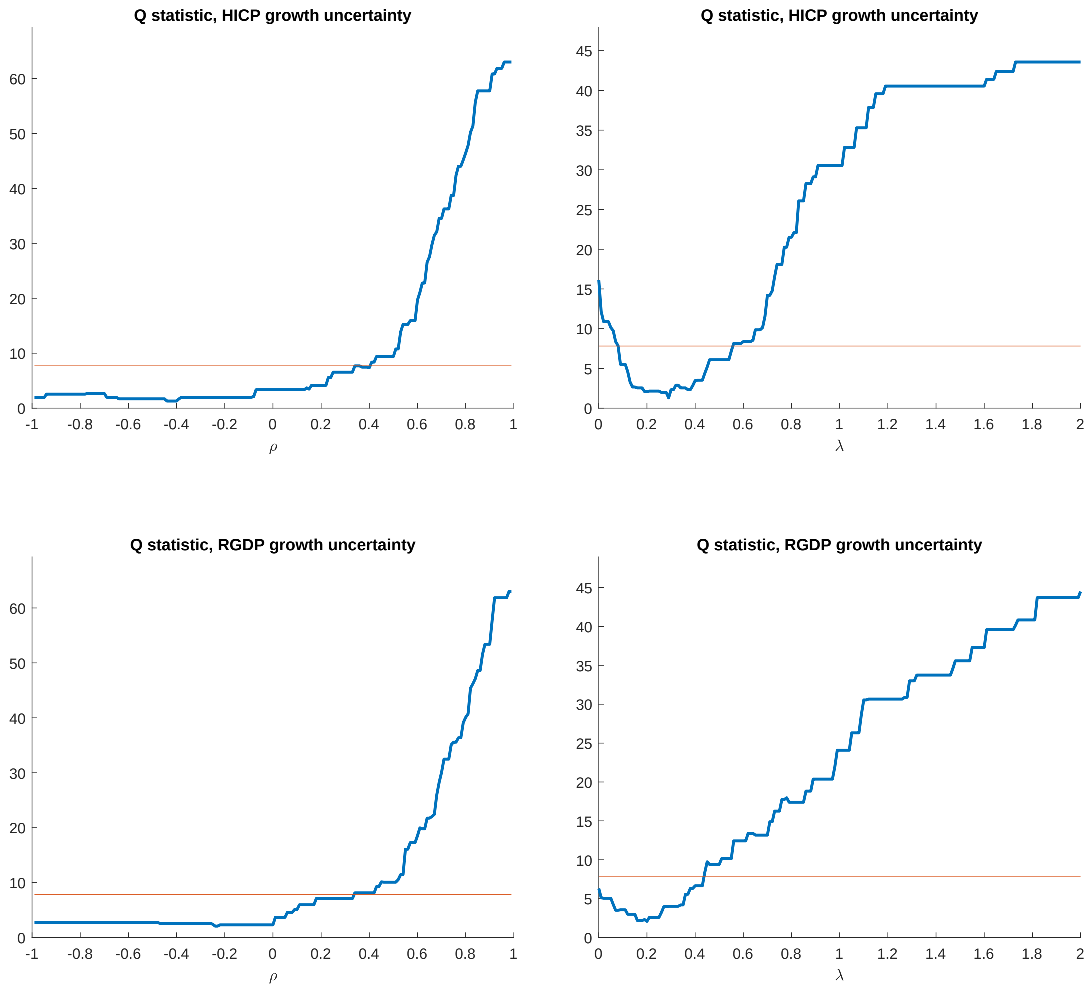


Figure 4.16: Q depending on ρ (left panels) and λ (right panels) for ECB SPF HICP inflation (upper panels) and real GDP growth (lower panels) without the restriction $\rho \geq 0$ imposed. The Q-values for ρ are obtained by setting $\lambda = \lambda^*$, and the Q-values for λ by setting $\rho = \rho^*$. The red line indicates the 5% critical value of the chi-squared distribution with 3 degrees of freedom, which Q approximately follows.

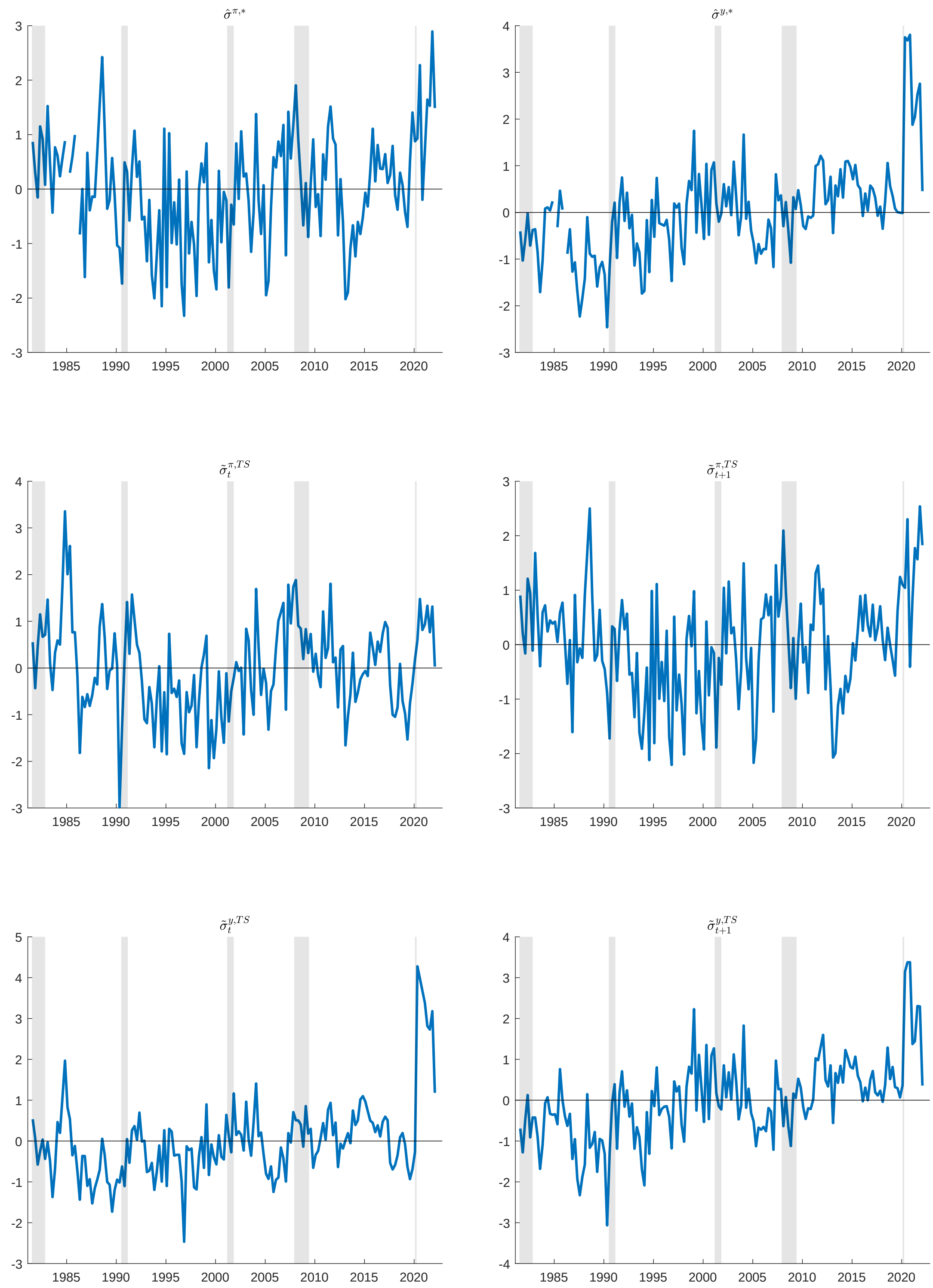


Figure 4.17: Standardized ex-ante uncertainty measures of the US SPF: FH inflation uncertainty (top left), FH growth uncertainty (top right). CY (middle left) and NY (middle right) inflation uncertainty (TS), CY (bottom left), and NY (bottom right) growth uncertainty (TS). For the abbreviations and descriptions of the individual uncertainty measures, see Table 4.8.

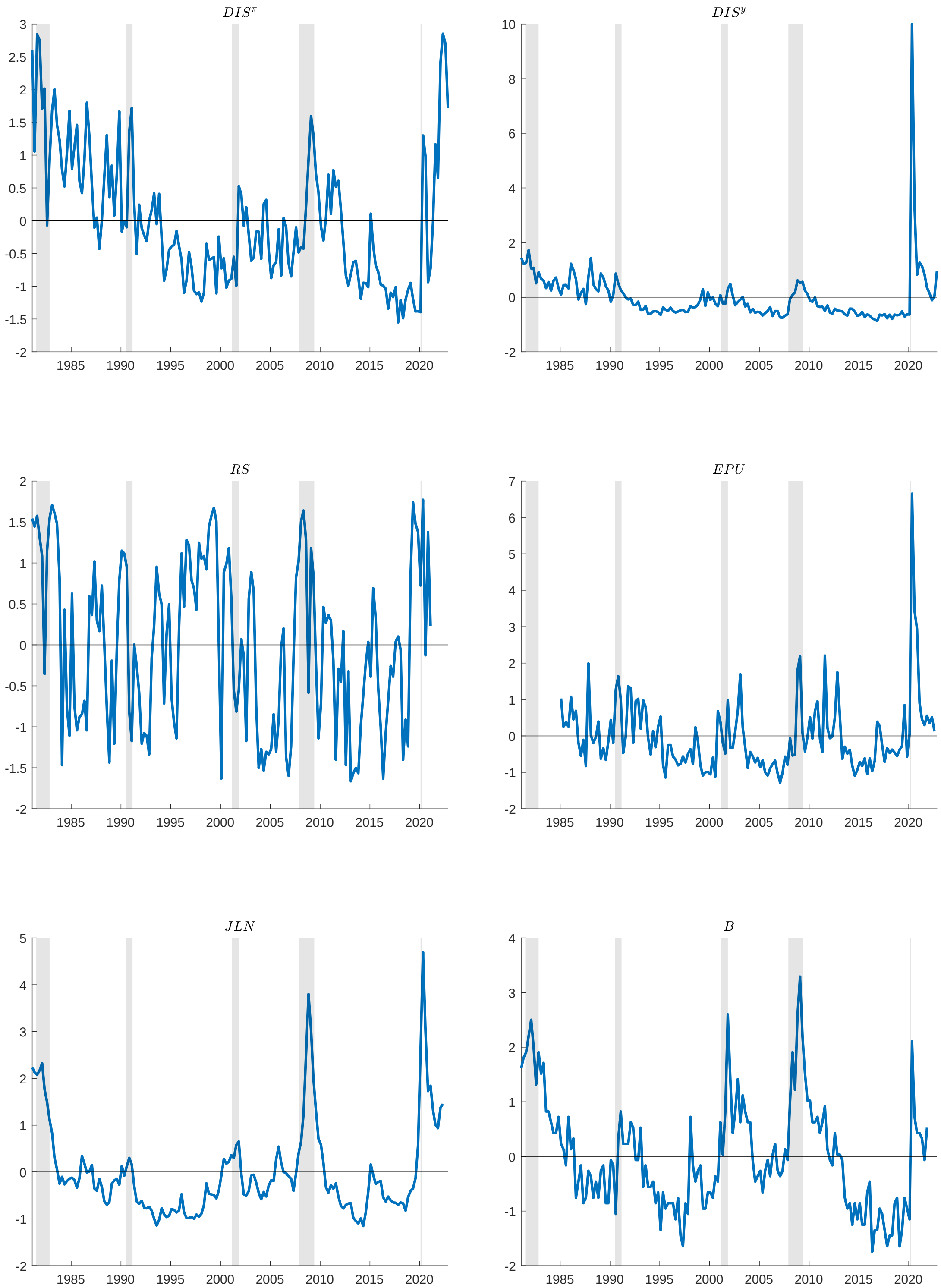


Figure 4.18: Standardized uncertainty measures: inflation disagreement (top left) and growth disagreement (top right), ex post uncertainty (middle left), Economic Policy Uncertainty (middle right), Macro Uncertainty (bottom left), rounding-based inflation uncertainty (bottom right). For the abbreviations and descriptions of the individual uncertainty measures, see Table 4.8.

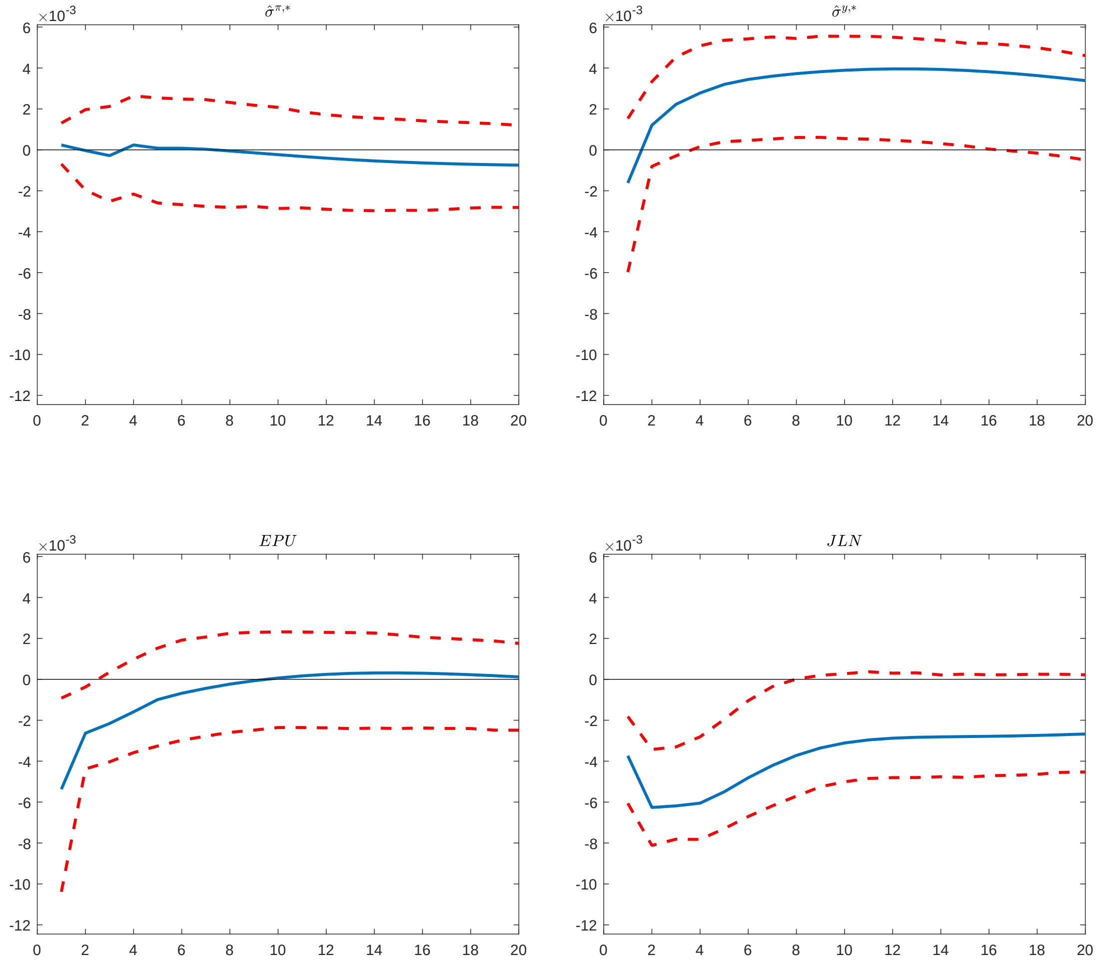


Figure 4.19: Responses of real GDP to a one-standard-deviation uncertainty shock in quarter 1. The red dashed lines are lower and upper 95% confidence bounds for each quarter estimated by bootstrapping. The uncertainty measures are FH inflation uncertainty (top left), FH growth uncertainty (top right), Economic Policy Uncertainty (bottom left), and Macro Uncertainty (bottom right). For the abbreviations and descriptions of the individual uncertainty measures, see Table 4.8. Estimation sample is 1985q1 to 2022q1.

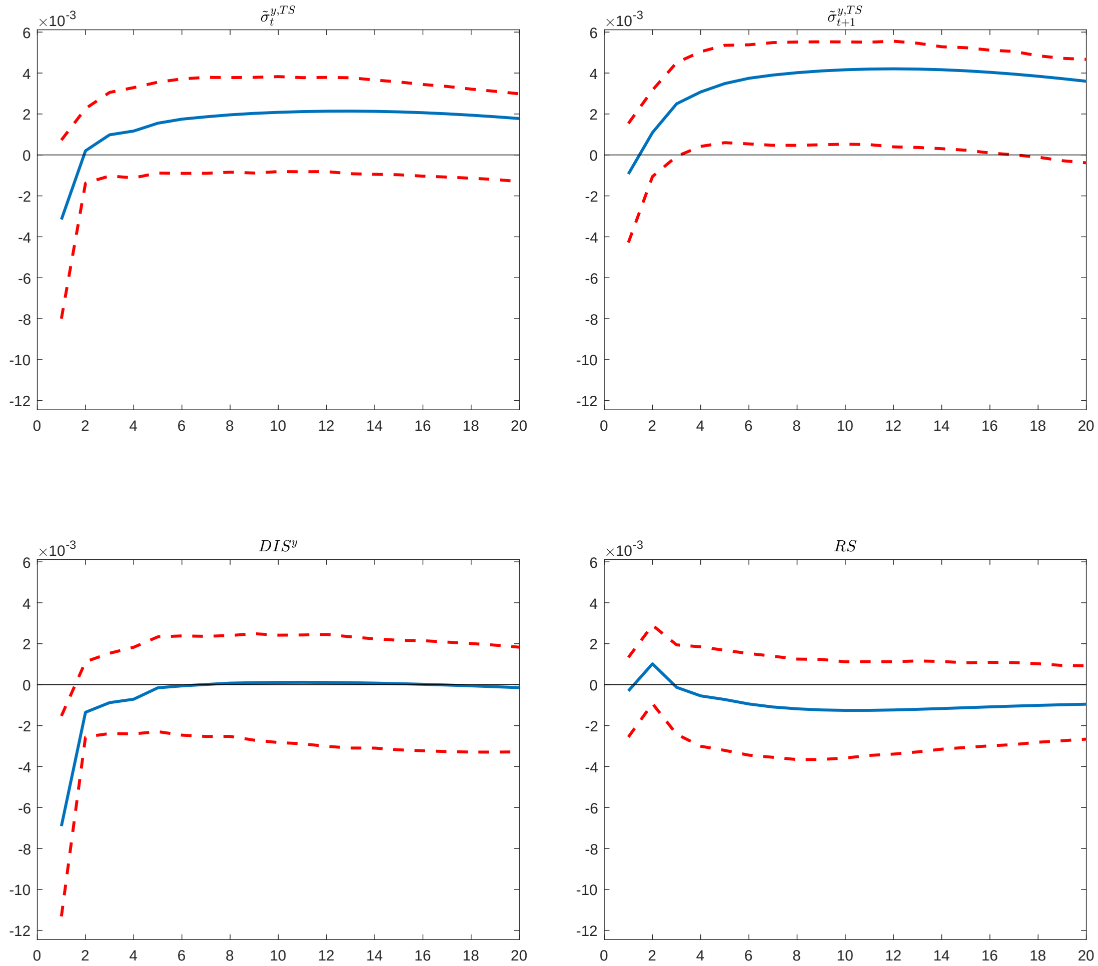


Figure 4.20: Responses of real GDP to a one-standard-deviation uncertainty shock in quarter 1. The red dashed lines are lower and upper 95% confidence bounds for each quarter estimated by bootstrapping. The uncertainty measures are CY (top left) and NY (top right) growth uncertainty, growth disagreement (bottom left), and ex post uncertainty (bottom right). For the abbreviations and descriptions of the individual uncertainty measures, see Table 4.8. The estimation sample is 1985q1 to 2022q1.

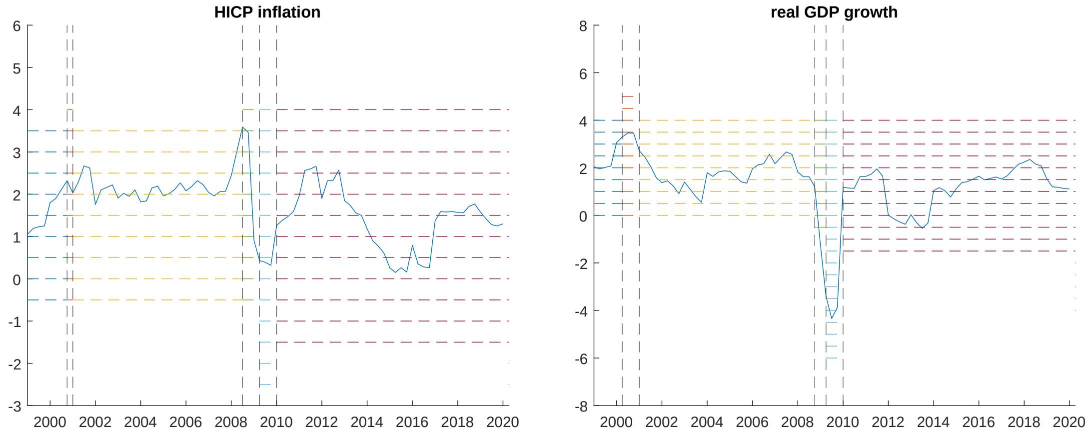


Figure 4.21: HICP (left panel) and real GDP (right panel) current-year growth forecasts together with histogram bins in the ECB SPF. Black dashed lines mark points with changes in the bin definitions in the ECB SPF. The graph is based on presentation slides for study of Clark et al. (2022).

5 Would Households Understand Average Inflation Targeting?[1]

5.1 Introduction

Central banks increasingly emphasize the importance of communicating their policies to the general public. A key goal is to steer the inflation expectations of households and firms such that real interest rates stabilize the economy even in periods when the policy rate is constrained. As the first major central bank, the U.S. Federal Reserve (Fed) announced in August 2020 that it would pursue a new monetary strategy labeled as 'flexible average inflation targeting' (AIT). Under this new strategy, the Fed aims to deliver inflation that averages 2% in the medium term. This implies that periods during which inflation has undershot the 2% target will be followed by periods of higher than 2% inflation. While such history-dependent monetary strategies have been shown to have good stabilization properties in theory, an open question is whether they will be successful in practice.

In this paper, we study whether households understand the implications of such a new monetary strategy and update their expected inflation paths accordingly. We do so by eliciting probabilistic expectations about the medium- and longer-term inflation outlook from 9,000 respondents from the Bundesbank Online Panel-Households (BOP-HH). Inflation expectations are well-anchored around the inflation aim of (close to but below) 2% which the European Central Bank (ECB) targeted until recently. We use randomized control trials to assess whether individuals would understand the implications of a hypothetical shift to an average inflation targeting strategy as the one introduced by

[1]Originally published at *Journal of Monetary Economics*, Vol. 129 (2022), S52-S66, `https://doi.org/10.1016/j.jmoneco.2022.02.006`.

the Fed in 2020. We also assess the quantitative adjustment of inflation expectations towards the target from below and above and study to what extent trust in the ECB affects the adjustment of inflation expectations. Importantly, we fielded our survey questions before the European Central Bank concluded its own framework review in July 2021.[2]

Our analysis shows that individuals asked to assume a hypothetical shift to average inflation targeting significantly increase their medium- and longer-term inflation expectations. Providing them with additional information about near-term inflation, respondents update their expected inflation paths in line with economic theory: they raise their inflation expectations if near-term inflation is below target and lower them when it is above target. The latter effect is quantitatively smaller, however, suggesting that individuals understand the asymmetry embedded in the Fed's communication which we follow in our information treatment. Individuals with high levels of trust in the ECB to deliver on its price stability mandate adjust their inflation expectations substantially more strongly. At the same time, respondents with the lowest levels of trust revise their inflation expectations downwards when provided with information about the alternative strategy.

While the stabilization properties of history-dependent monetary policy regimes are well understood in theory (see, e.g. Reifschneider and Williams (2000), Mertens and Williams (2019), and Amano et al. (2020)), surprisingly little is known about the key question our paper tries to answer: *Would households understand average inflation targeting?* To the best of our knowledge, the only other paper that aims to shed light on this question is Coibion et al. (2021). Using a daily online survey of U.S. individuals, these authors study whether the August 2020 introduction of the Fed's new monetary policy strategy had an impact on households' inflation expectations. They show that the announcement remained largely unnoticed by the general public. But even households who heard about the announcement did not incorporate the regime change into their

[2]The ECB started a review of its monetary policy strategy in January 2020, discussing a range of issues including an update of its definition of price stability. The introduction of some form of average inflation targeting seemed a realistic strategic option. For instance, in a speech at the ECB and its Watchers Conference 2020, President Christine Lagarde highlighted that, within the strategic review, it will be assessed *whether central banks should commit to explicitly make up for inflation misses when they have spent quite some time below their inflation goals.*

expectations. Moreover, when providing individuals with specific information about inflation targeting (IT) and AIT in a randomized control trial, Coibion et al. (2021) find that both treatments lead to similarly large and significant *reductions* in expected inflation, expected output growth and expected growth in personal income. Importantly, they also document that even one year after the Fed announcement, U.S. households had not incorporated the change of monetary strategy into their expectation formation.

At the surface, the findings in Coibion et al. (2021) seem to be in contrast with our results which show that individuals who are asked to assume an alternative monetary strategy akin to AIT on average *increase* their inflation expectations. While our survey design is inspired by their analysis, there are several important differences. First, our analysis studies a hypothetical rather than an actual shift of central bank policy as in Coibion et al. (2021). Second, the distribution of expected inflation is more tightly centered around the 2% inflation target in our survey of German individuals relative to comparable surveys of U.S. households where a larger share of households report inflation expectations considerably above the target. As a result, when provided with information about Fed policy, U.S. households tend to reduce their inflation expectations (Coibion et al., 2022b). Third, and most importantly, our results show that German households with low levels of trust in the ECB also *reduce* their inflation expectations when provided with information about the alternative monetary strategy, in line with the findings in Coibion et al. (2021) for the average treatment effect. Hence, differences across populations with respect to trust in the central bank may contribute to the observed differences in the effects of communicating monetary strategies.

At first sight, the increase in expected inflation under the hypothetical average inflation targeting regime appears quantitatively small. It ranges between 12 and 23 basis points at the two-to-three-year horizon across specifications. To assess the economic significance of this increase, we use a small-scale Dynamic Stochastic General Equilibrium (DSGE) model. We consider two model economies, where in one the central bank follows an inflation targeting regime, while in the other the central bank pursues an average inflation targeting policy. We calibrate the differences in medium-term inflation expectations to match the survey evidence and then simulate the two economies to compare their stabilization properties. We find that the volatility of inflation is substantially smaller under AIT compared to IT while the variability of the output gap is similar across regimes. Moreover, the frequency of hitting the lower bound of interest rates is considerably reduced under AIT. Combined, our results suggest that a history-dependent monetary strategy may have good stabilization properties to the extent that it is well communicated and understood by the public.

The paper is organized as follows. In Section 5.2, we describe our survey experiments and the collected data. Section 5.3 provides the results of our randomized control trial. The section also discusses to what extent households incorporate their inflation expectations into their consumption plans. The role of trust in the central bank for the adjustment of inflation expectations is analyzed in Section 5.4. Section 5.5 assesses the economic significance of the observed differences between IT and AIT using a small-scale DSGE model. Section 5.6 concludes.

5.2 Survey Experiments On Average Inflation Targeting

In this section, we describe the design of the randomized control trial (RCT) which we used to elicit household inflation expectations under different assumptions about the monetary regime and the near-term inflation outlook. We first provide some general information about the Bundesbank Online Panel-Households in which we conduct our experiments. We then discuss the details of the experimental setup which is inspired by the analysis in Coibion et al. (2021).

5.2.1 The Bundesbank Online Panel-Households

The survey experiments presented in this study were performed within the Bundesbank Online Panel-Households (BOP-HH). BOP-HH is a survey conducted at a monthly frequency to elicit consumer expectations about both macroeconomic and household-specific outcomes. The survey is representative of the German online population and has participants from the age 16 years and above. It contains a core set of general interest questions and typically includes a set of additional questions for the investigation of specific topics.[3]

[3] The German-language questionnaires and English-language translations are provided under `https://www.bundesbank.de/en/bundesbank/research/survey-on-consumer-expectations/`.

In our analysis, we use BOP-HH data collected in the October 2020 (Wave 10) with 1,903 respondents, and in January and February 2021 (Waves 13 and 14) with 2,342 and 4,737 respondents, respectively. In October 2020, we surveyed participants about their inflation expectations 2-3 years ahead, which we consider the horizon that best reflects the medium term. In January 2021, participants were asked about inflation expectations 5-10 years ahead, a horizon which we refer to as the longer term. We used the February 2021 wave with its larger sample size to construct two subsamples of about 3,000 and 1,700 respondents who were asked about the medium- and longer-term inflation expectations, respectively. While the survey has a small panel component, in our study we focus on results from three cross-sections of individuals. We ensure that our results are not driven by learning of the panel households across survey waves.[4] In addition to the questions regarding inflation expectations we also surveyed households about their spending intentions with respect to durable goods and their trust in the ECB's ability to achieve price stability.

5.2.2 Randomized Control Trial Set-Up

The basic AIT experiment conducted in October 2020, January, and February 2021 waves of the BOP-HH is a randomized control trial with a simple three-stage procedure, as summarized in Table 5.1. In the first stage, all participants receive general information about two monetary policy regimes. One is the strategy to aim at inflation rates close to, but below 2% in the medium term, as pursued by the ECB until July 2021.[5] We refer to this as the 'current strategy' or 'IT' in the remainder of the paper. The other monetary policy regime that we describe represents average inflation targeting as introduced by the U.S. Federal Reserve in August 2020. Specifically, we inform survey participants that the Fed now aims to steer the inflation rate at 2% on average. We describe in simple terms that this would involve that if inflation were to fall below 2% for some time, then the Fed

[4]There is a small number of about 300 individuals who remain in the panel between October 2020, the first wave in which we ran or experiment, and February 2021 which was the last wave. We confirm that our results are essentially unchanged when excluding these individuals from the analysis of the January and February waves.

[5]There is a marginal variation in the questionnaire between Wave 10 and 14, where Wave 10 swapped the "close to" formulation for a precise point target formulation. The comparison of the different waves shows that our results are robust to the communication variation.

Table 5.1: BOP-HH randomized control trial on average inflation targeting.

Stage 1	Infobox for all participants: ECB's current strategy aims at inflation rates close to, but below 2% in the medium term. An alternative strategy, as currently practiced by the Fed, to steer the inflation rate at 2% on average. Example that if inflation runs below the target, Fed will raise inflation above the target for some time.
Stage 2	All participants — assuming ECB is pursuing current strategy — are asked to assign probabilities for inflation 2-3 years (5-10 years)$^{\sharp:}$ ahead being . . . less or equal than 1% . . . greater than 1%, but at most 2% . . . greater than 2%, but at most 3% . . . greater than 3%
Stage 3	Participants are randomly sampled into one of five subgroups, A, B, C, D$^{\sharp:}$ or E$^{\sharp:}$, facing different assumptions about monetary policy and current inflation. Then, participants are asked again to assign probabilities as in Stage 2: Group A — alternative strategy Group B — current strategy; 2021 inflation at 1% Group C — alternative strategy; 2021 inflation at 1% Group D — current strategy; 2021 inflation at 3%$^{\sharp:}$ Group E — alternative strategy; 2021 inflation at 3%$^{\sharp:}$

Note: The RCT as described was conducted in October 2020 to elicit medium-term (2-3 years ahead) expectations, and in January to elicit longer-term (5-10 years) expectations, both Waves using subsamples A, B, and C. Subsamples D and E in Stage 3 of the RCT, marked with a $^{\sharp:}$, are exclusive to Wave 14 of February 2021 where both medium- and longer-term expectations were elicited, so Stage 3 comprises subsamples A, B, C, D and E.

would aim to raise inflation above 2% for some time thereafter. That is, we explicitly adopt the asymmetric communication used by Federal Reserve Chairman Powell in his August 2020 Jackson Hole announcement.[6] We choose this formulation as it was geared towards a broader public and appears straightforward to understand. We refrain from using the term "average inflation targeting" as a potential strategy the ECB could adopt as it was only vaguely discussed by policymakers in the euro area. Nonetheless, for brevity, we refer to this 'alternative strategy' as 'AIT' in the remainder of the paper.

[6] Note that the Fed communication of the new strategy explicitly did not mathematically define a reaction function, but purposefully embedded some flexibility.

In the second stage, participants are asked to make a probabilistic assessment for inflation 2-3 years (October 2020 and February 2021) and 5-10 years (January and February 2021) ahead. For simplicity, we offer respondents a simple four-bin histogram where they are asked to assign probabilities to the events that inflation over the solicited horizon is less or equal 1%, greater than 1% but at most 2%, greater than 2% but at most 3%, or greater than 3%.

In the third stage of the experiment, participants are randomly sampled into five subgroups and are again asked for their probabilistic assessment of inflation 2-3 or 5-10 years ahead. In two of the groups, we ask respondents to continue assuming that the ECB follows its current strategy but to make additional assumptions about near-term inflation. Specifically, one of the groups is asked to consider that inflation in the calendar year 2021 (the beginning of the forecast horizon) would be at 1% and the other group that 2021 inflation would equal 3%. The purpose of this additional assumption is to investigate individuals' expected inflation *path* back to the target. In a third group, we ask participants to assume that the ECB would be pursuing the alternative strategy described before, without further assumptions about 2021 inflation. Finally, the last two groups are asked to assume that the ECB would be following the alternative strategy under the additional assumption that inflation in 2021 would be at 1% or 3%, respectively.

5.3 Inflation Expectations under Different Monetary Regimes

This section provides an analysis of expected inflation over the medium and longer run elicited under the two different monetary policy regimes. As the February 2021 wave with about 4,700 respondents allows us to provide the most granular analysis, we focus on this wave. We provide comparisons with the other waves along the way to document the robustness of our findings.

5.3.1 Expected Inflation Outcomes in the Medium and Longer Run

Figure 5.1 compares the average probability distribution of medium (left panel) and longer-term (right) inflation expectations under the two different regimes, without additional

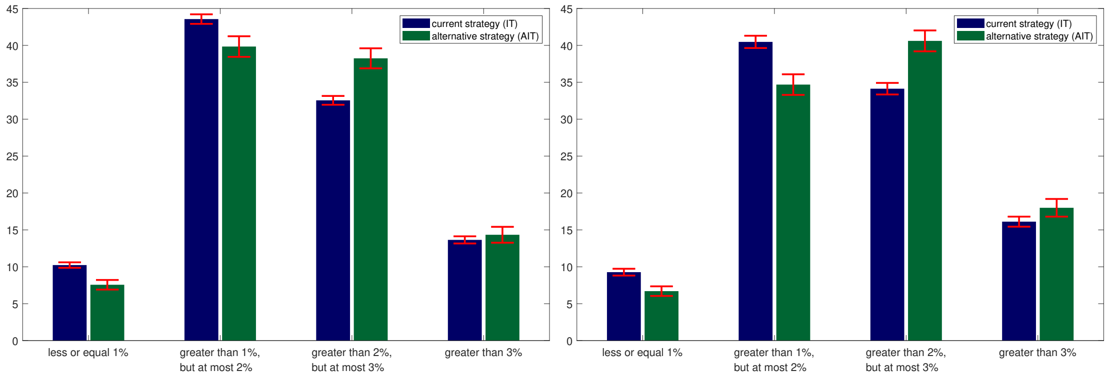


Figure 5.1: Left panel: Inflation expectations for 2-3 years ahead elicited in February 2021 (Wave 14) of BOP-HH. Right panel: Inflation expectations for 5-10 years ahead elicited from a disjoint sample of respondents in February 2021 (Wave 14) of BOP-HH. Dark blue bars show average subjective probabilities of medium-term inflation with respondents assuming current monetary policy (IT). Dark green bars show the average subjective probabilities of respondents assuming the alternative strategy (AIT). A two-standard error band is plotted in red. Test statistics are provided in Table 5.5 in the Appendix.

assumptions about near-term inflation. Dark blue bars show inflation expectations when respondents are asked to assume that the ECB is pursuing its current price stability objective to aim at an inflation rate close to, but below 2% over the medium term. The bars of the second and third bin taken together show that for both forecast horizons, more than 75% of the mass is assigned to inflation falling between 1 and 3%. Hence, both medium and longer-term inflation expectations of German households appear to be relatively well-centered around the ECB's price stability objective of 2% at the time when the survey was conducted. The average subjective probability of inflation exceeding 3% is roughly 13-15% while that of inflation falling below 1% is about 10% for both horizons.

The dark green bars show average subjective probabilities when respondents are asked to assume that the ECB would pursue the alternative strategy. At both forecast horizons, the reported probabilities for inflation below 2% are considerably lower than under the current strategy. In contrast, the probability mass is shifting towards higher expected inflation. Statistically, these average subjective probabilities under the two strategies are different, bin by bin and jointly. The p values of the corresponding t test statistics and for the $\mathcal{T}^2$ test statistic of Hotelling (1931) are shown in Table 5.5 in the Appendix. A two-standard error bar is plotted in red to visualize the sampling uncertainty around the mean differences for every bin. Despite the upward shift of inflation expectations under AIT, the distribution is a bit more symmetric than under IT with about 75-80% of the probability mass allocated to inflation being between 1 and 3% at both horizons. The February 2021 results are consistent with the findings of October 2020 for the medium run, as shown in the left panel of Figure 5.5 in the Appendix. As realized inflation had increased markedly over this time period, our results thus seem to be robust to changes in the underlying inflation dynamics.

To summarize, this first analysis suggests that inflation expectations for the medium and longer run would be somewhat higher under average inflation targeting as compared to inflation targeting. At first sight, it might appear surprising that longer-term inflation expectations are also shifted up significantly under the assumed AIT strategy. The reason is that under AIT the central bank is supposed to bring inflation back to the 2% target in the medium term, such that well-anchored longer-term inflation expectations should be at target. We can offer three potential explanations for this finding. First, households may consider 5-10 years into the future as the medium term, in contrast to most central bankers who would associate the medium term with shorter horizons of 2-3 years. Second, households may well understand the asymmetry implicit in the communication by the Fed chair which we use in our information treatment. Accordingly, while they understand that inflation may rise above 2% for some time after a period of below-target inflation, they may not anticipate inflation to be pushed *below* 2% after a period of higher-than-target inflation. Third, assuming an arguably implausible level of sophistication of households, they may understand that average inflation in the longer term will be higher under AIT as the economy will hit the lower bound on interest rates less often under such a policy regime.

5.3.2 Comparing Mean Inflation Expectations

In the previous section, we have shown that the distribution of medium- and longer-term inflation expectations is significantly different across monetary policy regimes, with more probability mass at inflation levels above 2% under AIT. In this section, we translate these probabilistic assessments into mean inflation expectations to measure the magnitude of the increase in expected inflation.

We back out mean inflation expectations as the first moment of the individual respondents' histograms broadly following Engelberg et al. (2009). The resulting distributions of mean inflation expectations for the medium term across the two policy regimes are provided in Figure 5.6 in the Appendix. They confirm the pronounced upward shift under the hypothetical AIT strategy as compared to the IT strategy. In particular, while the mass of mean inflation expectations falling in the interval 1.5 to 2% is reduced moving from IT to AIT, the share of mean expectations falling into the 2 to 2.5% and the 2.5 to 3% bin increase markedly. The average of the mean inflation expectations across individuals shifts from 2.03% to 2.17% in the October 2020 wave and from 2.02% to 2.15% in the February 2021 wave. A similar shift can be observed for the longer-term mean inflation expectations, elicited in January 2021 and February 2021, as shown in Figure 5.7 in the Appendix.

5.3.3 Adjustment of Inflation Expectations towards the Target

Thus far we have established that households have somewhat higher medium-term and longer-term inflation expectations under AIT than under IT. A key aspect of make-up strategies such as AIT is that inflation expectations work as "automatic stabilizers" in the sense that they increase when inflation is running below target and decline when inflation is running above target for some time. Here, we assess the quantitative adjustment of inflation expectations toward the target from below and above. As noted above, the current communication of the Fed with respect to their flexible AIT strategy is asymmetric. While it is clearly stated that the Fed would try to steer inflation above the 2% target after it has been undershooting the target for some time, the opposite is not explicitly mentioned. We follow this communication and provide participants with an explanation of the alternative (AIT) strategy in line with the Fed's asymmetric statement.

To assess the adjustment of expected inflation paths, we asked subgroups of respondents in Stage 3 of the RCT to assume that the ECB would continue to follow its current (IT) strategy and, in addition, to assume that 2021 inflation would be at 1%, that is, below

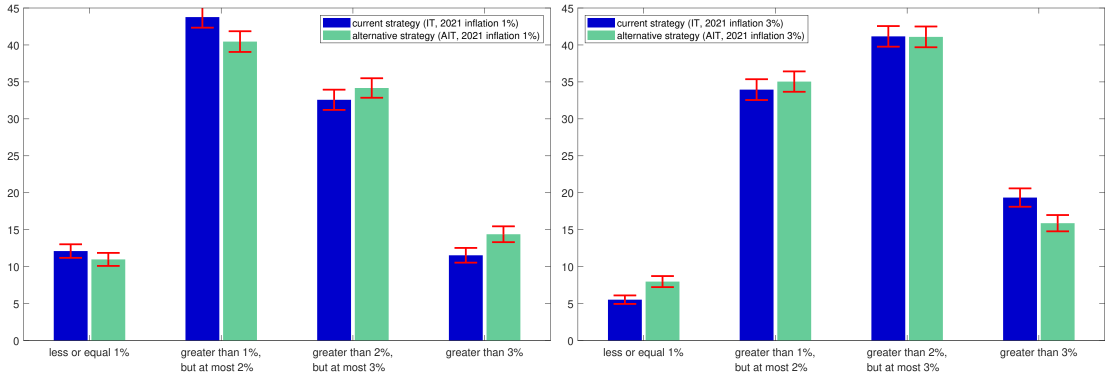


Figure 5.2: Inflation expectations for 2-3 years with 2021 inflation at 1% and 3%, respectively (February 2021 Wave 14). Left panel: Respondents are asked to assume 2021 inflation at 1%. Right panel: Respondents are asked to assume 2021 inflation at 3%. Inflation expectations for 2-3 years ahead elicited in February 2021 (Wave 14) of BOP-HH. Royal blue bars show average subjective probabilities of medium-term inflation with respondents assuming current monetary policy (IT). Teal bars show the average subjective probabilities of respondents assuming the alternative strategy (AIT). A two-standard error band is plotted in red. Test statistics are provided in Table 5.5.

the inflation target. Another group was asked to assume that 2021 inflation would be at 3%, i.e. above the inflation target. Analogously the two groups received information about the alternative strategy.

The left-hand panel of Figure 5.2 compares the outcomes for the 1% treatment, the right-hand panel those for the 3% treatment. With the additional assumption of inflation being 1% in 2021, respondents under AIT report significantly less probability mass in the below-target bins than the respondents under IT, but instead more mass than the IT group in the above-target bins. This is in line with the desired adjustment of inflation expectations. The effect is to some extent reversed under the 3% assumption treatment. While IT and AIT respondents attribute about the same probability mass to medium-term inflation falling in the 1 to 2% and the 2 to 3% bins, there is a significant difference in the outer bins. Survey participants asked to assume the alternative strategy predict a larger probability that inflation will fall below 1% than IT respondents; the opposite is true for inflation above 3%. Hence, households appear to understand that a lower (higher) than target near-term inflation rate is more likely to be followed by an above (below) target medium-term inflation rate under AIT than under IT.

To assess the quantitative magnitude of these effects, we again consider mean inflation expectations derived from the individual respondents' probabilistic assessments. The cross-sectional distributions of mean expectations for the medium-term under the various treatments are provided in Figure 5.8 in the Appendix. While the distributions have fairly similar shapes under the two regimes, the average expected mean inflation is about 10 basis points higher for AIT than for IT respondents when asked to assume inflation in 2021 to be at 1%, but is about 10 basis points lower for AIT than IT respondents under the 3% treatment. Again, these results suggest that households perceive a stronger adjustment of inflation towards the target under AIT as compared to the IT strategy.

We now assess the statistical significance of the shifts in mean inflation expectations. Specifically, we follow Coibion et al. (2019) and compute the differences between the individual post-treatment mean inflation expectation and the pre-treatment mean inflation expectation formed under the assumption of current monetary policy. We then regress these differences on dummy variables indicating whether or not an individual was sampled into a specific subgroup. The estimation equation is then simply

$$\text{mean}_i^s - \text{mean}_i^{IT} = \sum_s^S \delta_s d_{s,i} + u_i,\ s \in \{AIT, IT1\%, AIT1\%, IT3\%, AIT3\%\}. \quad (1)$$

The subscript i represents the individual respondent in the cross-section, and sub- and superscript s reflects the sub-sample, corresponding to a specific combination of monetary policy strategy and 2021 inflation the participant is asked to assume. The coefficient δ_s captures the change in medium-term inflation expectations due to treatment s, measured in basis points. The variable $d_{s,i}$ is the treatment dummy. All mean differences are pre-post comparisons, calculated with respect to the information that the ECB is currently pursuing its inflation-targeting strategy, such that $\text{mean}_i^{IT} \equiv \text{mean}_i^{pre}$ and $\text{mean}_i^s \equiv \text{mean}_i^{post}$.

Table 5.2 reports the results. The first column shows the treatment effects for medium-term inflation expectations. The coefficient for AIT is 13 basis points and statistically significant at the 1% level. Hence, the medium-term inflation expectations of respondents asked to assume the alternative strategy are significantly higher without any additional assumptions about near-term inflation. In contrast, the coefficient for IT1% is a strongly

Table 5.2: Baseline regression results for mean inflation expectations 2-3 years and 5-10 years ahead (February 2021 Wave 14)

	$\text{mean}_i^s - \text{mean}_i^{IT} = \sum_s^S \delta_s d_{s,i} + u_i,\ s \in \{AIT, IT1\%, AIT1\%, IT3\%, AIT3\%\}$			
	Expectations 2-3 years ahead		Expectations 5-10 years ahead	
	(1)	(2)	(3)	(4)
AIT	0.13^{***}	0.23^{***}	0.12^{***}	0.21^{***}
	(0.03)	(0.05)	(0.03)	(0.05)
IT1%	-0.09^{***}	-0.14^{***}	0.02	0.03
	(0.03)	(0.05)	(0.03)	(0.04)
AIT1%	0.03	0.05	0.07^{**}	0.10^{**}
	(0.03)	(0.05)	(0.03)	(0.04)
IT3%	0.25^{***}	0.37^{***}		
	(0.03)	(0.05)		
AIT3%	0.28^{***}	0.42^{***}		
	(0.03)	(0.05)		
AIT1% − IT1%	0.12^{**}	0.19^{***}	0.05	0.07
AIT3% − IT3%	0.03	0.05		
Observations	2970	1848	1745	1057
Adjusters only	No	Yes	No	Yes

Note: Sub- and superscript s refers to a certain monetary policy strategy with specific assumptions about near-term inflation. For instance, $s = AIT3\%$ denotes 'average inflation targeting', with respondents additionally asked to assume that the inflation rate in 2021 would be at 3%. Columns (1) and (3) report the change in inflation expectations 2-3 (5-10) years ahead for all respondents, while columns (2) and (4) show results only for individuals who updated their inflation expectations after the information treatment for the respective horizons. All regressions use weighted survey responses where the weights ensure the representativeness of the sample. Asterisks ($^{***}, ^{**}, ^{*}$) indicate significance levels of 1, 5 and 10% for the usual t test statistics for coefficient estimates and for the F test statistics for differences in coefficient estimates.

statistically significant negative 9 basis points, suggesting individuals lower their medium-term inflation expectations under the current regime when prompted to assume near-term inflation below target. This suggests that below-target inflation rates might become entrenched in expectations under the IT strategy. This is not true, however, for the participants asked to assume that the ECB follows the alternative (AIT) strategy. These individuals slightly increase their medium-term inflation expectations, albeit not statistically significant. Turning to the coefficients for the *IT3%* and *AIT3%* treatments, we see that both groups significantly increase their medium-term inflation expectations by similar magnitudes. While the estimated coefficients are interesting for themselves, it is instructive to compare the differences across regimes, given a specific assumption about start-year inflation. The corresponding F- test results are shown in the last two rows for the 1% and 3% groups, respectively. Mean inflation expectations increase by a statistically significant 12 basis points more under AIT as compared to IT when asked to assume below-target near-term inflation. They also increase somewhat more under AIT under the assumption of above-target near-term inflation, but that difference is not statistically significant.

These results capture all survey responses, including those of individuals who did not adjust their perceived inflation outcomes after the information treatment. To evaluate the magnitude of the adjustments for those who revise their expectations, the second column of Table 5.2 repeats these results only for those individuals who changed their assessment in response to the treatment. The share of non-adjusters is sizeable, with roughly 30% of respondents not changing their expectations between Stage 2 and Stage 3 of the experiment. Nonetheless, the vast majority of remaining participants, dubbed 'adjusters', seems to understand the key mechanism of AIT. Inflation expectations increase by a highly statistically significant 23 basis points for households asked to assume the alternative strategy. Individuals in the IT1% group lower their inflation expectations by a significant 14 basis points while those in the *AIT1%* group increase them by 5 basis points. The difference of 19 basis points is highly statistically significant, showing households would understand the concept of inflation overshooting when it is currently running below target. In contrast, as shown by the positive but insignificant difference between the *AIT3%* and the *IT3%* groups, they do not anticipate an undershooting of inflation when it is currently running above the target. As discussed above, this response is in line with the asymmetric Fed communication that we treat individuals with.

The third and fourth column of Table 5.2 report the regression results for longer-term inflation expectations. Due to the smaller number of respondents in this treatment arm, we only compare the two strategies without additional assumptions about near-term inflation and under the 1% assumption. The coefficients show that households raise their longer-term inflation expectations by similar magnitudes as their medium-term inflation expectations when asked to assume the alternative monetary strategy: 12 basis points for all households and 21 basis points considering only those who adjusted their assessment after the treatment. Interestingly, in contrast to the medium-term inflation expectations, households asked to assume the current strategy and 2021 inflation at 1% did not lower their longer-term inflation expectations. This suggests that they understand that inflation would move back to target over the longer-run under the IT strategy. Again in contrast to the medium-term expectations, individuals asked to assume the alternative strategy and 2021 inflation at 1% significantly raised their 5-10 years ahead inflation expectations. This might imply that they perceive this longer horizon as representative of the "medium term" used in the description of the two strategies that we provide to the survey participants.

Table 5.6 in the Appendix shows that these results are robust to various modifications in the regression specification. In particular, they are largely unchanged when we use unweighted as opposed to weighted individual survey responses as in our baseline regression. The weights ensure that the survey responses are representative of the German online population in terms of a range of socioeconomic characteristics including gender, age, education status, income, region of domicile, etc. Our results are also robust with respect to possible outliers or heteroskedasticity. Therefore, throughout the remainder of the paper, we show results for the weighted data and employ simple OLS estimations.

A potential concern with our analysis based on repeated cross-sections is that inflation expectations could respond to changes in actual inflation from wave to wave. Indeed, annualized consumer price inflation in Germany increased substantially from below two percent in the fall of 2020 to above three percent in early 2021. Several factors contributed to this shift: a temporary reduction in the VAT rate from 19% to 16% between July 2020 and December 2020, an increase in carbon taxes on fuel and other energy sources, and base effects related to the strong decline of commodity and other prices at the beginning of the COVID-19 pandemic. Table 5.7 in the Appendix compares the regression results for the three different survey waves conducted in October 2020, January and February 2021. While the coefficients vary somewhat in magnitude, the main conclusions are unchanged: households understand the concept of average inflation targeting and adjust their inflation expectations in the desired way.

5.3.4 Adjustment of Consumption Plans

So far, we have shown that German households adjust their inflation expectations in line with the theoretical prescription of average inflation targeting. But even if central banks are able to steer inflation expectations in the desired way, the stabilization properties of history-dependent monetary policy rules also depend on the impact that movements in expected inflation have on aggregate demand and in particular, household consumption. Despite the tight link between inflation expectations and consumption via the real interest rate channel in economic theory, prior research e.g. by Bachmann et al. (2015) has found at best mixed evidence supporting such a connection.

We follow Bachmann et al. (2015) and evaluate the relation between expected inflation and intended durable consumption. To this end, we complemented our RCT with two follow-up questions. The first asks whether or not it is currently a good time to buy durable goods. The second lets respondents state a reason for the answer to the previous question. The augmented survey design is shown in Table 5.8 in the Appendix. We then specify a probit model regressing the binary variable capturing the reported consumption intention on the individual mean inflation expectations post-treatment and a set of demographic controls. The average marginal effects implied by both regressions are shown in Table 5.9 in the Appendix. In both specifications, the impact of inflation expectations on households' spending intentions is essentially zero, confirming the findings of Bachmann et al. (2015) in our data.[7] While inflation expectations do not play a role, several socio-demographic characteristics significantly affect individuals' spending intentions. In particular, wealthier and more educated individuals show a higher propensity to consume durables, consistent with Armantier et al. (2015). The same is true for female respondents, who seem to be more sensitive to price developments, as also pointed out by D'Acunto et al. (2021a) and D'Acunto et al. (2021b). In contrast, individuals who have lived in East Germany prior to 1990 feature a strongly lower propensity to consume durables. Our results also show that the missing link between inflation expectations and reported spending plans is not driven by the weighting of respondents to match the online population of German individuals. We also do not find a link between inflation

[7] We also do not find a significant impact on durable spending intentions when using the difference in mean inflation expectations before and after the treatment as an explanatory variable.

uncertainty extracted from the probabilistic inflation expectations and spending plans, in contrast to the results reported in Coibion et al. (2019). Finally, there is no statistically significant link between (changes in) inflation expectations and consumption plans when excluding participants who do not adjust their expectations in response to the provided information.

Despite the missing link between inflation expectations and durable spending intentions, we find that households *qualitatively* associate expected price changes with consumption. After deciding whether now is a 'good time to buy' or 'not a good time to buy', respondents were asked to specify a reason for their answer.[8] As Table 5.10 in the Appendix shows, expected price increases for prices in general and for those of durable goods are the major reported reasons for respondents to answer that currently is a good time to buy durables. This is true both prior to and after the treatment and across the two monetary policy regimes. When participants were asked to give a reason why currently it is not a good time to buy, smaller price increases only play a marginal role. The primary reason chosen for not planning to purchase durables is the lack of need for replacement, followed by a precautionary savings motive where participants do not want to spend savings.

In sum, the information treatment ignites little change in respondents' reported consumption plans. We interpret these findings as evidence that the magnitude of the shift in expected inflation induced by a hypothetical change in monetary policy may be too small to meaningfully affect private households' spending intentions. In fact, as pointed out by Andrade et al. (2020), household consumption is more likely to respond to broad changes in the inflation regime rather than to small variations in inflation.

5.4 The Role of Trust for the Adjustment of Inflation Expectations

It is well understood by central bankers and academic economists that central banks' ability to steer the inflation expectations of households and firms crucially depends on

[8] Answering options were provided as a list, with the ordering of the entries randomized. Alternatively, participants could type a reason into a free-form text field, but this option was rarely used in practice. The reasons that could be chosen by the respondents and the resulting distribution of answers are shown in Table 5.10 in the Appendix. The top panel lists the reasons chosen at Stage 2, prior to treatment. The bottom panel shows the selected reasons post-treatment.

their credibility (Blinder, 2000). This is particularly important for AIT which makes monetary policy history-dependent by promising to off-set past misses of inflation in the future (Amano et al., 2020). In this section, we analyze whether the observed differences between AIT and IT are possibly amplified by the degree of central bank credibility.

In the survey literature, central bank credibility is often associated with high reported levels of trust in the central bank, see e.g. Christelis et al. (2020). They find that individuals with higher trust in the ECB on average have lower inflation expectations and report lower uncertainty about future inflation. To investigate whether trust in the central bank affects households' inflation expectations under different monetary policy regimes, we explicitly asked participants to what extent they trust the ECB's ability to achieve price stability in the February 2021 wave of the BOP-HH survey. Respondents could choose any value between '0 = Do not trust at all' and '10 = Trust entirely', and could also exit stating 'I do not know the ECB'. The modal value of trust across respondents is 5, with a mean of 4.7. Few reported very high levels of trust and about 10 percent of respondents did not show any trust in the ECB. Importantly, we elicited these trust values *before* the information treatment that is part of our RCT experiment.[9]

To assess the extent to which trust in the central bank affects the mean inflation expectations of private households across monetary policy regimes, we interact the treatment dummy with the trust variable as follows:

$$\text{mean}_i^s - \text{mean}_i^{IT} = \sum_s^S \delta_s d_{s,i} + \sum_s^S \gamma_s d_{s,i} \times \text{Trust}_i + u_i, \tag{2}$$

where $s \in \{AIT, IT1\%, AIT1\%, IT3\%, AIT3\%\}$ denotes the various treatment arms. Due to the interaction terms in the regression model, the treatment effect of a particular monetary policy strategy s now also depends on the level of trust in the ECB, so that the treatment effect under s is given by $\delta_s + \gamma_s \times \text{Trust}_i$.

The results are shown in Table 5.3. We test the Null hypotheses that the shifts of mean expected inflation under the various information treatments are unchanged once we interact with trust. We evaluate the corresponding F-tests at the 10%, median, and the 90% quantile of the trust distribution. Focusing first on the medium-run

[9] The distribution of reported trust values is shown in Figure 5.9 in the Appendix.

inflation expectations shown for all respondents and for adjusters only in the first and second column, we make the following observations. First, without providing additional assumptions about near-term inflation, both the coefficient on the treatment dummy *AIT* and on the interaction with trust are highly significant. Interestingly, the former switches sign and becomes negative while the latter is positive. This indicates that individuals with higher trust in the ECB increase their inflation expectations more strongly under AIT, while individuals with low trust revise their inflation expectations downward. As a result, the medium-term inflation expectations are negative at the 10th percentile, but positive at the median and strongly positive at 40 bps at the 90th percentile of the trust distribution. This is somewhat in contrast to Coibion et al. (2021) who find that the average respondent in a survey of U.S. individuals adjusts inflation expectations negatively when confronted with information about the new Fed strategy.

Second, under the additional assumption that current inflation would initially be below target at 1%, the difference between AIT and IT becomes more pronounced at medium and high trust levels relative to the baseline specification. For low levels of trust, the difference between mean inflation expectations is still positive, but statistically insignificant. Individuals with intermediate and high levels of trust, however, revise their inflation expectations more strongly upwards. Third, respondents with low and intermediate levels of trust also somewhat raise their inflation expectations relative to IT when near-term inflation is assumed to be 3%. This adjustment is negative only for high levels of trust, in line with the expected make-up effect that medium-term inflation expectations under AIT should be lower than under IT. As in the baseline specification without taking trust into account, these differences are largely statistically insignificant. Importantly, the difference between AIT and IT declines as trust increases, suggesting that confidence in the ECB to achieve its price stability objective contributes to the adjustment of inflation expectations in line with the central banks' intentions, independently of the monetary regime. For the longer run, displayed in columns 3 and 4 of the table, the differences between the mean inflation expectations are positive and highly statistically significant for intermediate and high levels of trust. Under the additional 1% assumption they are not statistically significant at any level of trust.[10]

[10] While we only show results based on weighted data in Table 5.3 with the weights ensuring that the survey responses are representative of the German online population, our findings are robust to using

Table 5.3: Inflation expectations and trust in the central bank (February 2021 Wave 14)

$$\text{mean}_i^s - \text{mean}_i^{IT} = \sum_s^S \delta_s d_{s,i} + \sum_s^S \gamma_s d_{s,i} \times \text{Trust}_i + u_i,\ s \in \{AIT, IT1\%, AIT1\%, IT3\%, AIT3\%\}$$

	Expectations 2 - 3 years ahead		Expectations 5 - 10 years ahead	
	(1)	(2)	(3)	(4)
AIT	-0.24^{***}	-0.41^{***}	-0.01	0.01
	(0.07)	(0.12)	(0.06)	(0.10)
IT1%	-0.15^{**}	-0.26^{**}	0.04	0.07
	(0.07)	(0.11)	(0.06)	(0.10)
IT3%	0.07	0.15		
	(0.07)	(0.12)		
AIT1%	-0.09	-0.14	0.07	0.11
	(0.07)	(0.11)	(0.06)	(0.10)
AIT3%	0.24^{***}	0.38^{***}		
	(0.06)	(0.10)		
AIT × Trust	0.08^{***}	0.13^{***}	0.03^{**}	0.04^{**}
	(0.01)	(0.02)	(0.01)	(0.02)
IT1% × Trust	0.01	0.02	0.00	-0.01
	(0.01)	(0.02)	(0.01)	(0.02)
IT3% × Trust	0.04^{***}	0.04^{**}		
	(0.01)	(0.02)		
AIT1% × Trust	0.03^{**}	0.04^{**}	0.00	0.00
	(0.01)	(0.02)	(0.01)	(0.02)
AIT3% × Trust	0.01	0.01		
	(0.01)	(0.02)		
AIT				
at 10%-Quantile	-0.16^{***}	-0.15^{**}	-0.01	0.05
at 50%-Quantile	0.16^{***}	0.24^{***}	0.14^{***}	0.21^{***}
at 90%-Quantile	0.40^{***}	0.63^{***}	0.23^{***}	0.33^{***}
AIT1% − IT1%				
at 10%-Quantile	0.08	0.16	0.03	0.03
at 50%-Quantile	0.16^{***}	0.22^{***}	0.03	0.09
at 90%-Quantile	0.22^{**}	0.28^{**}	0.03	0.12
AIT3% − IT3%				
at 10%-Quantile	0.14^{*}	0.17		
at 50%-Quantile	0.02	0.08		
at 90%-Quantile	-0.07	-0.01		
Observations	2957	1841	1735	1050
Adjusters only	No	Yes	No	Yes

Note: Columns (1) and (3) report the change of inflation expectations 2-3 (5-10) years ahead for all respondents, while columns (2) and (4) show results only for individuals who updated their inflation expectations after the information treatment for the respective horizons. Asterisks (***, **, *) indicate significance levels of 1, 5 and 10% for the usual t test statistics for coefficient estimates and for the F test statistics for the differences between *IT1%* versus *AIT1%* and *IT3%* versus *AIT3%*, respectively. As the monetary policy regimes are now interacted with the trust level, the F tests are set up to assess the Null of the equality of a given IT-AIT pair of partial derivatives, calculated for the 10%-, the 50%- and 90%-quantiles of the trust distribution. For example, the Null of the F test for the comparison $AIT1\% - IT1\%$ is obtained as $H_0 : \hat{\delta}_{AIT1\%} - \hat{\delta}_{IT1\%} + (\hat{\gamma}_{AIT1\%} - \hat{\gamma}_{IT1\%}) \times \text{Trust}_{50\%} = 0$. All regressions use weighted survey responses where the weights ensure the representativeness of the sample.

Figure 5.3 visualizes the effect of trust in the ECB on the adjustment of inflation expectations. The left column shows results for all respondents, corresponding to the regression results in the first column of Table 5.3, while the right column provides results looking only at the roughly two-thirds of adjusters who have changed their inflation expectations after the information treatment (corresponding to the second column of the table).

The top row shows results for the AIT treatment without additional information about start-year inflation. For each level of trust, we plot the change in inflation expectations from before to after the information treatment as green circles. The red diamonds show the average of these observations for each level of trust. The regression line in the top-left chart plots the estimated relation between the change in inflation expectations and trust.[11] The line is upward-sloping, showing that higher levels of trust are associated with a stronger adjustment of inflation expectations to the provided information. More importantly, however, it highlights that for levels of trust below 3, the adjustment of inflation expectations tends to be negative. This is in line with the baseline results of Coibion et al. (2021) who document a negative average impact of information about the new monetary strategy of the Federal Reserve on U.S. households' inflation expectations. While these authors don't report trust in the Fed, their result of a negative impact on the provided information on inflation expectations might be consistent with our results if many of their respondents had low levels of trust. The corresponding chart in the top-right panel of Figure 5.3 shows the same regression for adjusters only. While the impact of trust on the adjustment of inflation expectations is a bit more pronounced, the conclusion that individuals with levels of trust below 3 tend to lower their inflation expectations remains valid.

The middle and bottom panels of Figure 5.3 show the corresponding relationships between trust and the adjustment of inflation expectations for the groups of respondents provided with the additional 1% and 3% assumption, respectively. Rather than plotting all individual changes in inflation expectations, for ease of exposition, we simply show

Huber (1981) weights or unweighted data. They are also robust with respect to possible outliers or heteroskedasticity.

[11] That is, the intercept corresponds to the coefficient on the AIT treatment dummy (−0.24% for all respondents), and the slope corresponds to the coefficient on the interaction with trust (0.08%).

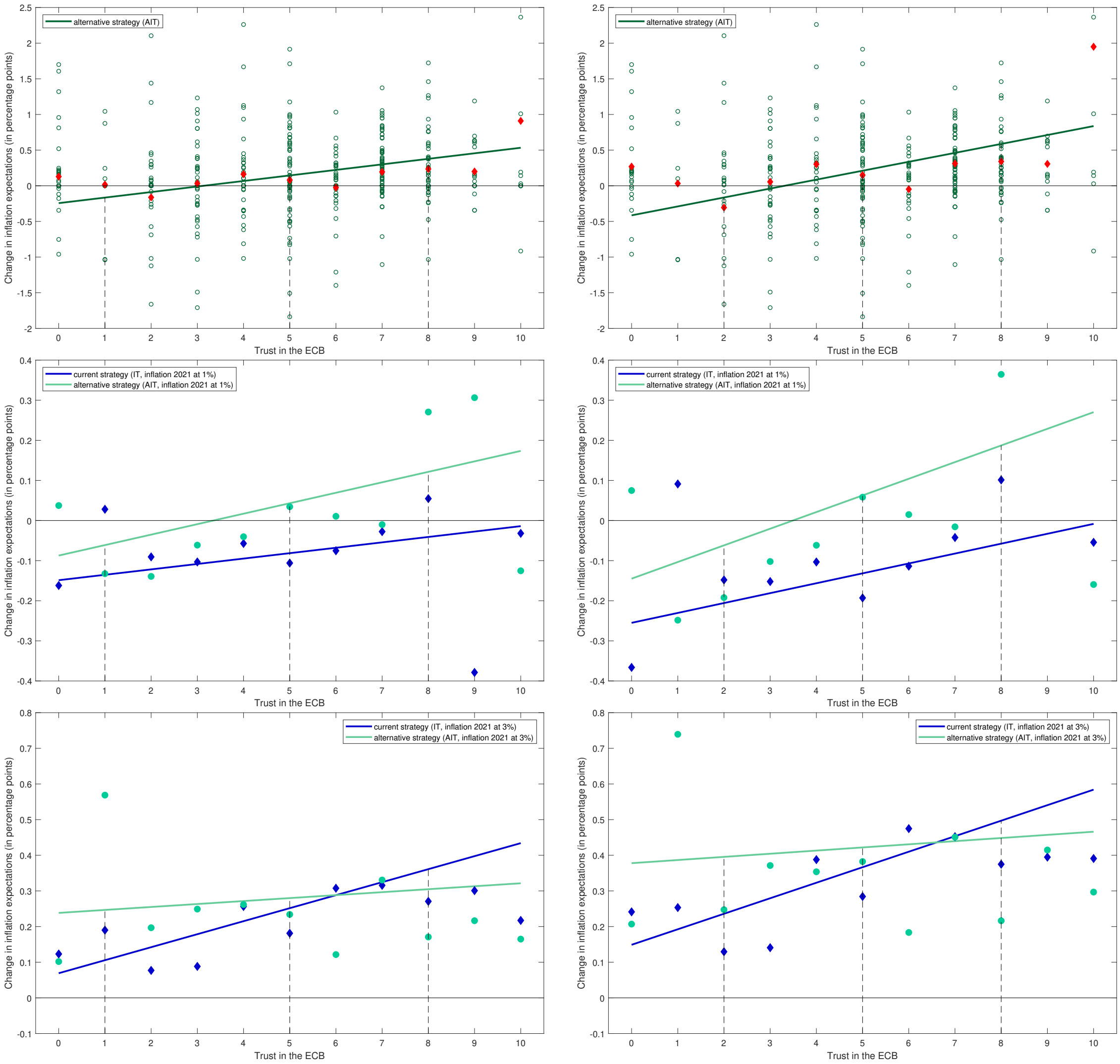


Figure 5.3: Change in inflation expectations when controlling for the level of trust in the ECB's ability to achieve price stability. Note: Left side: All respondents in the respective treatment group. Right side: Only respondents that have adjusted their inflation expectations post-treatment (adjusters only). Top panel: Observations exceeding -2 and 2.5 are not displayed. Observations depicted in red are group means computed for each level of trust. The middle and bottom panel report only group means for different trust levels. Dashed lines represent 10%-, 50%- and 90%-quantiles. The regression lines plot the marginal impact of trust on the change in inflation expectations for a given information treatment.

the average per treatment arm and level of trust (green for the alternative (AIT) strategy and blue for the current (IT) strategy). The regression lines correspond to the coefficients based on the individual responses documented in Table 5.3. For the groups asked to assume below-target start-year inflation the green line is steeper and consistently above the blue, indicating that trust in the ECB amplifies the adjustment of inflation expectations. As for the AIT treatment with additional information, only AIT1% respondents with levels of trust below 3 reduce their inflation expectations. That said, for all levels of trust in the ECB IT1% respondents lower their inflation expectations, consistent with the notion that low initial inflation would not lead to a subsequent inflation overshoot under the inflation targeting policy.

A slightly different picture emerges from the bottom panel showing the groups of respondents asked to assume a 3% start-year inflation. Here, individuals treated with information about the alternative strategy consistently revise their inflation expectations upwards, and the impact of trust on these revisions is fairly low. In contrast, and somewhat counterintuitively, the respondents asked to assume the current IT policy increase their inflation expectations more strongly the more they trust in the ECB's ability to achieve price stability.

In light of the important role of trust in the ECB it is instructive to study the sociodemographics of individuals with different levels of trust. Table 5.4 provides the percentage shares of various sociodemographic characteristics for individuals with low (trust ≤ 1), intermediate (trust $= 5$) and high levels of trust (trust ≥ 8), respectively. The pattern that emerges from these numbers is that respondents with low levels of trust tend to have fewer years of education and lower incomes. Regarding the age structure, individuals between 40 and 60 years have lower levels of trust, while those younger than 40 years tend to have higher levels of trust. The distribution of trust is relatively flat for respondents older than 60 years. Both employment status and gender also do not seem to play a role. That said, individuals who lived in East Germany before the wall came down tend to have lower levels of trust.

In sum, these results suggest that high trust in the ECB increases both the level of medium-term and longer-term inflation expectations and the strength of the adjustment of inflation expectations back towards target, especially when inflation starts from below.

Table 5.4: Percentage share breakdown of socio-economic characteristics of the Wave 14 sample, categorized by trust levels

	Low trust	Median trust	High trust
Age (in years)			
below 40	25.1	34.9	41.7
40 to 60	46.7	35.3	30.3
over 60	28.2	29.8	28.0
Female	48.7	60.1	43.7
HH income (in EUR)			
under 1500	16.3	11.6	10.9
1500 to 3000	39.6	37.2	28.1
3000 to 5000	32.8	34.0	40.6
over 5000	11.3	17.2	20.4
Employed	62.9	60.2	60.7
College degree	14.7	18.9	32.9
High school degree	22.9	30.3	51.1
East GER pre 1990	27.2	16.1	15.8
No. of observations	364	592	440

Note: Figures correspond to percentage shares of a total of 100% or each variable for the respective trust level, i.e., column-wise. *Low trust* numbers comprise individuals with trust levels of 0 and 1; *Median trust* with a trust level of 5; and *High trust* with trust levels of 8, 9 and 10, respectively.

Moreover, since trust is distributed quite unevenly among different socioeconomic groups, central bank communication efforts targeting constituents with low levels of trust might have particularly large benefits for central banks aiming to steer the public's inflation expectations.

5.5 Assessing the Economic Significance

In the previous sections, we have shown that households increase their inflation expectations when prompted to assume an alternative monetary strategy akin to average inflation targeting. While the increase is highly statistically significant, the magnitude of the observed shift in mean inflation expectations of about 12 to 23 basis points across the various specifications might appear small at first glance. In this section we gauge the economic significance of this increase in a simple New Keynesian model. We first summarize the model's aggregate demand and supply conditions in equilibrium, as well as the monetary policy rules corresponding to AIT and IT, respectively. We then calibrate the model to match the estimated differences in medium-term inflation expectations and simulate the model to compare the stabilization properties of the two monetary regimes.

5.5.1 The Model's Equilibrium Conditions

The equilibrium conditions are log-linearized around the non-stochastic steady state, such that $\overline{X}$ reflects the steady state of variable X, and $x_t = \log(X_t) - \log(\overline{X})$ represents its log-linearized form.

The benefits of AIT and other history-dependent monetary policy rules derive from their ability to steer inflation expectations. In practice, however, the expectation channel appears to be constrained as expectations only adjust sluggishly (e.g. Coibion et al. (2018b)) or as agents adjust their expectations only partially to policy announcements (e.g. Mauersberger and Nagel (2018)). To account for the empirical evidence, we mitigate the expectation channel in the model.[12] Specifically, we follow Galí et al. (2007) and Bilbiie (2019) and assume that a share of households λ is limited to borrow or save in financial markets, which yields the following aggregate demand equation.

$$x_t = E_t\left[x_{t+1}\right] - \frac{1-\lambda}{\sigma(1-\lambda\left(1+\psi\right))}\left(r_t - E_t\left[\pi_{t+1}\right] - r_t^n\right). \tag{3}$$

A higher λ, i.e. a higher share of households subject to a borrowing or lending constraint, attenuates the real interest rate channel and, hence, the role of inflation expectations, $E_t\left[\pi_{t+1}\right]$, on aggregate demand today. We set $\lambda = 0.2$ to match our survey evidence. This reflects our finding that when asked about whether it is currently a good time to buy durable goods, around 20% of households replied they would not adjust their spending decisions in response to expected price changes because of limited access to credit and or liquid savings (see Table 5.10 in the Appendix). The value of $\lambda = 0.2$ is also in line with empirical estimates based on DSGE models by Coenen and Straub (2005), Ratto et al. (2009), and Hoffmann et al. (2021a).

[12]For simplicity, we focus on the consequences of constraining the expectation channel from a static perspective only. For models that endogenize the interaction between monetary policy and expectation formation, see Melosi (2017), Falck et al. (2021), and Carvalho et al. (2023), forthcoming.

The variable r_t captures the policy rate. The difference between the policy rate and expected inflation, $r_t - E_t[\pi_{t+1}]$, is the real interest rate. The variable x_t reflects the output gap in deviations of output from its natural level. The natural rate of interest, $r_t^n = (1-\rho^r)\,\overline{r}_t^n + \rho^r r_{t-1}^n + \sigma^r \varepsilon_t^r$, is assumed to evolve exogenously. A drop in ε_t^r reflects the effects of a negative demand shock. The intertemporal elasticity of substitution equals $1/\sigma$. The labour supply elasticity equals ψ.

For the firm sector, we assume that some firms ω are backward-looking, similar to Galí and Gertler (1999). Rather than setting their prices optimally based on expectations about future inflation, these firms set prices by a rule of thumb. This assumption results in the Phillips curve

$$\pi_t - \pi^* = \beta\chi E_t[\pi_{t+1} - \pi^*] + \frac{\omega}{\theta}\chi(\pi_{t-1} - \pi^*) + \kappa x_t + u_t. \tag{4}$$

Firms set their prices around the central bank's inflation target, $\pi^* > 0$. The variable $u_t = \rho^u u_{t-1} + \sigma^u \varepsilon_t^u$ captures an exogenous supply disturbance in the form of a cost-push shock ε_t^u. The parameter β captures the discount factor. The parameter θ represents the fraction of forward-looking firms which cannot adjust their prices the next quarter. The weight on expected inflation in the Phillips curve, χ, as well as the slope of the Phillips curve, κ, are defined as

$$\chi = \frac{\theta}{\omega(1-\theta+\theta\beta)+\theta} \text{ and } \kappa = \frac{(\sigma+\psi)(1-\omega)(1-\theta)(1-\beta\theta)}{\omega(1-\theta+\theta\beta)+\theta}. \tag{5}$$

Both coefficients are decreasing in ω. Thus, the higher the share of firms which set their prices based on a rule of thumb, the weaker the effects of expected inflation on inflation today, and the flatter the slope of the Phillips curve. In our simulations we set the share of backward-looking firms to $\omega = 0.75$, in line with Galí and Gertler (1999). This also accounts for the apparent flattening of the Phillips curve, as discussed in Clarida (2019) and Bobeica et al. (2019), among others.

The model is closed by the respective monetary policy rule. To highlight the implications of make-up strategies on inflation and inflation expectations, we compare an AIT regime with a regime of IT, where the latter is currently common practice among major central banks.[13] Under IT, the policy rule equals

$$r_t = \max\left\{0, \overline{r}^n + \pi^* + \phi^\pi\left(\pi_t - \pi^*\right) + \phi^x x_t\right\}, \tag{6}$$

where the central bank stabilizes *current inflation* around its inflation target π^* and the output gap. Under AIT, the policy rule is instead defined as

$$r_t = \max\{0, \overline{r}^n + \pi^* + \frac{\phi^\pi_{AIT}}{1+l}\sum_{k=0}^{l}\left(\pi_{t-k} - \pi^*\right) + \phi^x x_t\}, \tag{7}$$

with l denoting the number of lags of inflation. Now, the central bank focuses on stabilizing *average inflation* around its inflation target over the past l periods. Similarly to Amano et al. (2020) we set $l = 8$ in our simulation exercise below.[14] To make the monetary policy responses to demand and supply disturbances comparable between the two monetary policy regimes, we match the volatilities of the policy rates. This implies $\phi^\pi_{\text{AIT}} > \phi^\pi$ in our simulation exercise.[15]

5.5.2 The Implications Of Demand And Supply Disturbances

In this section we assess the dynamics of both model economies by simulating random sequences of demand and supply disturbances over 100 quarters. We assume that both types of shocks are equally important. Crucially, we calibrate the difference in households' inflation expectations under AIT versus IT 2-3 years ahead to the value of 12 basis points obtained under the additional assumption of 2021 inflation being one percent below target, see Table 5.2.[16]

The first row of Figure 5.4 displays the annualized level of policy rates in one set of model simulations. The graph shows that the frequency of hitting the effective lower

[13]See Hammond et al. (2012) for a thorough overview of inflation targeting central banks.

[14]Amano et al. (2020) have shown that, in a model similar to ours, without an explicit inflation target a lag length of $l = 6$ resembles closely a policy regime of price-level targeting.

[15]To account for the effective lower bound, we are solving the model by using the OccBin-toolbox provided by Guerrieri and Iacoviello (2015).

[16]To do so, the initial value of inflation is set independently of supply and demand disturbances to 1% below the model's inflation target in both policy regimes. Then, the calibration is set to imply 12 basis points higher inflation expectations under AIT compared to IT. The parameter values are within the range of empirical estimates for the Euro area and Germany by Hoffmann et al. (2021a).

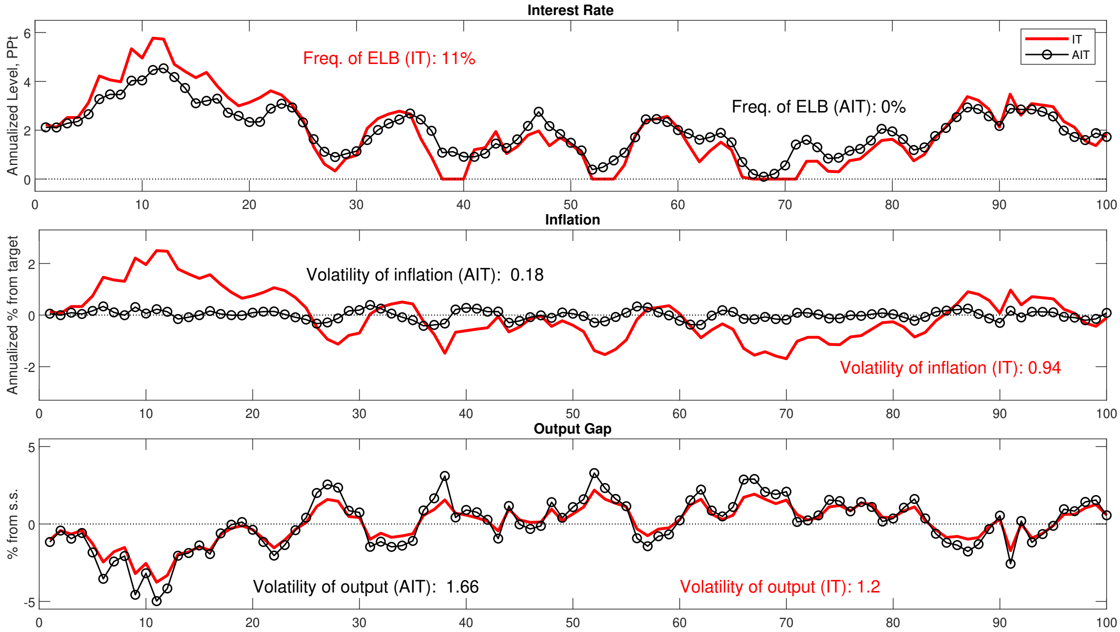


Figure 5.4: Movements of the interest rate, inflation and output. Note: Parameter settings: Annualized inflation target $\pi^* \times 400 = 2\%$, discount factor $\beta = 0.995$, intertemporal elasticity of substitution $1/\sigma = 1$, labour supply elasticity $\psi = 1$, the probability of adjusting prices is $1 - \theta = 0.3$. The share of backward-looking firms is $\omega = 0.75$, of constrained households is $\lambda = 0.2$. Taylor rule coefficients are $\phi^\pi = 1.5$ and $\phi^{\pi AIT} = (1 + l) \times \phi^\pi$. Persistence of demand and supply shocks are $\rho^r = 0.7$ and $\rho^u = 0.7$, respectively, with $\varepsilon_t^r = \varepsilon_t^u = .0006$.

bound (ELB) is significantly higher under an IT policy. Over our simulation horizon, the economy remains at the ELB for 11 quarters (i.e. 11% of the time) under an IT rule. In contrast, under the AIT rule the economy does not hit the ELB in our simulation.

The stabilization properties to inflation and output in the presence for this sequence of demand and supply shocks are provided in the second and third row of Figure 5.4. The volatility of inflation is about five times smaller under AIT compared to IT, while the variability of the output gap is relatively similar across the two monetary policy regimes.

The findings for this particular simulation are representative of the average dynamics in the two economies, repeating random sequences of demand and supply disturbances over 1000 times. Figures 5.10 and 5.11 in the Appendix confirm this also for the case where either demand or supply shocks would dominate. Moreover, Figure 5.12 shows that reducing the real interest rate channel by increasing the share of borrowing-constrained households λ by 100% also does not alter the main conclusion of this section. The model economy remains at the ELB less often and experiences lower inflation volatility under AIT, while the volatility of output is similar under the two monetary policy regimes. In sum, these results suggest that even small differences in expected inflation substantially enhance the stabilization properties of AIT relative to IT.

5.6 Conclusion

In this study, we have investigated whether private households would understand key differences in the characteristics of monetary policy strategies. We collected survey data from about 9,000 participants in the Bundesbank Online Panel-Households and treated respondents with different pieces of information regarding the ECB's hypothetical monetary policy regime in a randomized control trial. Our results show that the average respondents' probability mass is well-centered around the ECB's inflation aim of (close to, but below) 2% when assuming current monetary policy, but significantly increases when asked to assume an alternative monetary policy regime akin to average inflation targeting. Moreover, households appear to understand well the adjustment path of inflation back to the target. AIT households treated with a 1% start-year inflation assumption raise inflation expectations significantly more than the corresponding IT households, while a 3% start-year inflation assumption produces the opposite effect, albeit of a smaller magnitude.

The adjustment is particularly strong for individuals who, prior to the treatment, report intermediate to high levels of trust in the ECB to achieve it price stability objective. In contrast, individuals with low levels of trust reduce their inflation expectations when being informed about AIT. While expected price increases are in fact key drivers of spending intentions, the empirical link between expected inflation and durable consumption plans is fairly weak, however, and does not significantly differ across assumed monetary regimes. We assess the economic significance of our survey results in a small New Keynesian Model. We find that inflation is considerably less volatile and the frequency of hitting the lower bound of interest rates is strongly mitigated under AIT compared to IT.

In sum, these results suggest that — if communicated clearly — households seem to understand the intended mechanics of average inflation targeting and adjust their inflation expectations in line with theory. Moreover, as trust in the central bank appears to strengthen the effect of information about monetary strategies on inflation expectations, targeting people with low levels of trust may have particularly large benefits when communicating a new central bank strategy. An interesting direction for follow-up work is to assess the understanding of the new monetary strategy communicated by the ECB in July 2021, after our survey experiments. A first analysis by Hoffmann et al. (2021b) suggests that German households indeed raised their inflation expectations somewhat when provided with information about the new strategy and in particular when being informed that inflation might exceed the target for some time after periods of below-target inflation rates.

5.7 Appendix

5.7.1 Additional Figures

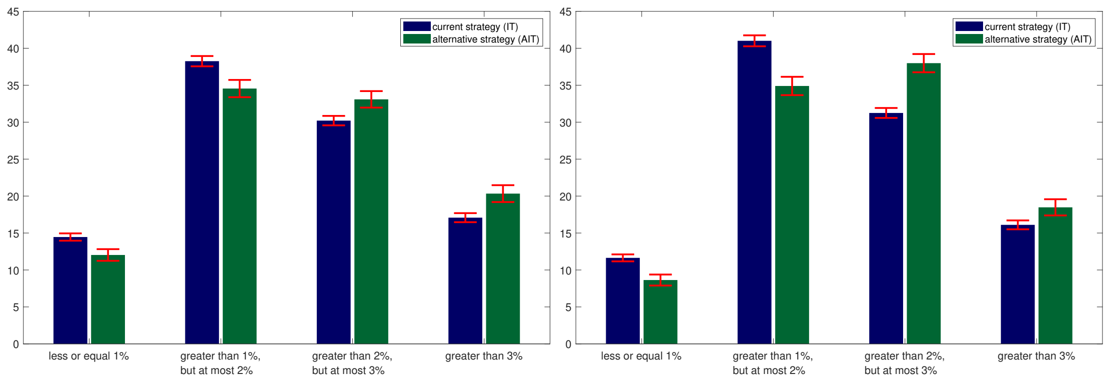


Figure 5.5: Inflation expectations for 2-3 years and 5-10 years ahead (October 2020 Wave 10 and January 2021 Wave 13). Left panel: Inflation expectations for 2-3 years ahead elicited in October 2020 (Wave 10) of BOP-HH. Right panel: Inflation expectations for 5-10 years ahead elicited in January 2021 (Wave 13) of BOP-HH. Dark blue bars show average subjective probabilities of inflation with respondents assuming current monetary policy (IT). Dark green bars show the average subjective probabilities of respondents assuming the alternative strategy (AIT). A two-standard error band is plotted in red. Test statistics are provided in Table 5.5.

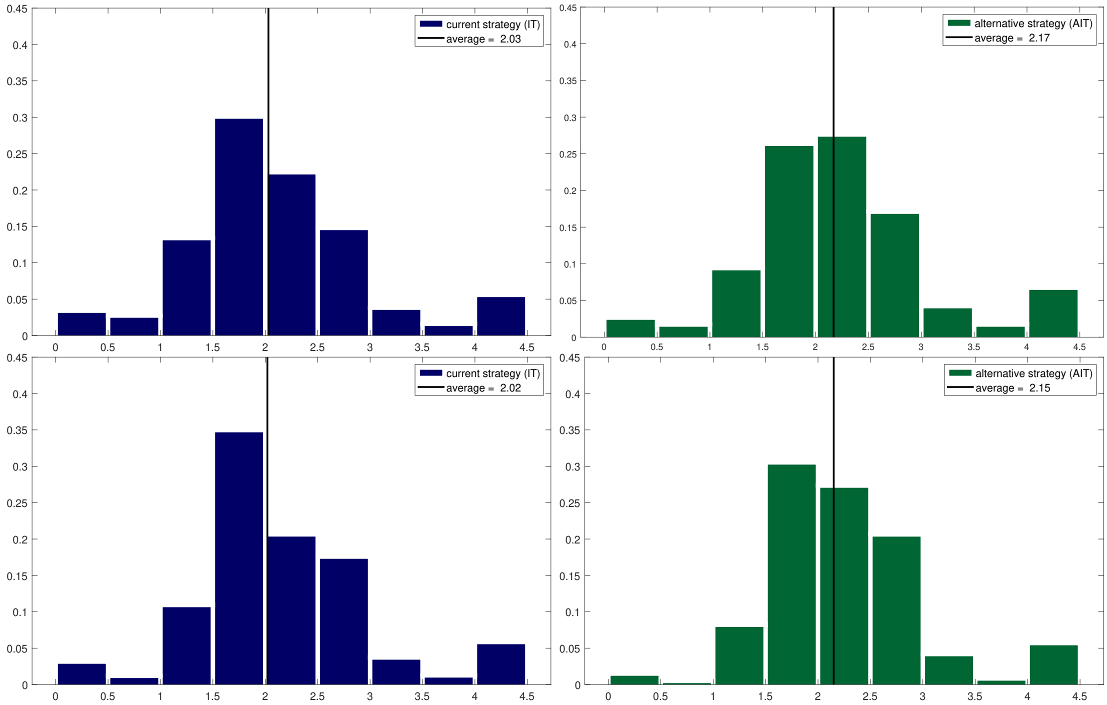


Figure 5.6: Mean inflation expectations 2-3 years ahead (October 2020 Wave 10 and February 2021 Wave 14): IT vs. AIT. Top row: October 2020 Wave 10 of BOP-HH. Bottom row: February 2021 Wave 14 of BOP-HH. Left panels: Dark blue bars show mean inflation expectations for the medium term with respondents assuming IT. Right panels: Dark green bars show mean inflation expectations for the medium term with respondents assuming AIT. An upright black line depicts the respective average of the mean inflation expectations. Test statistics are provided in Table 5.5.

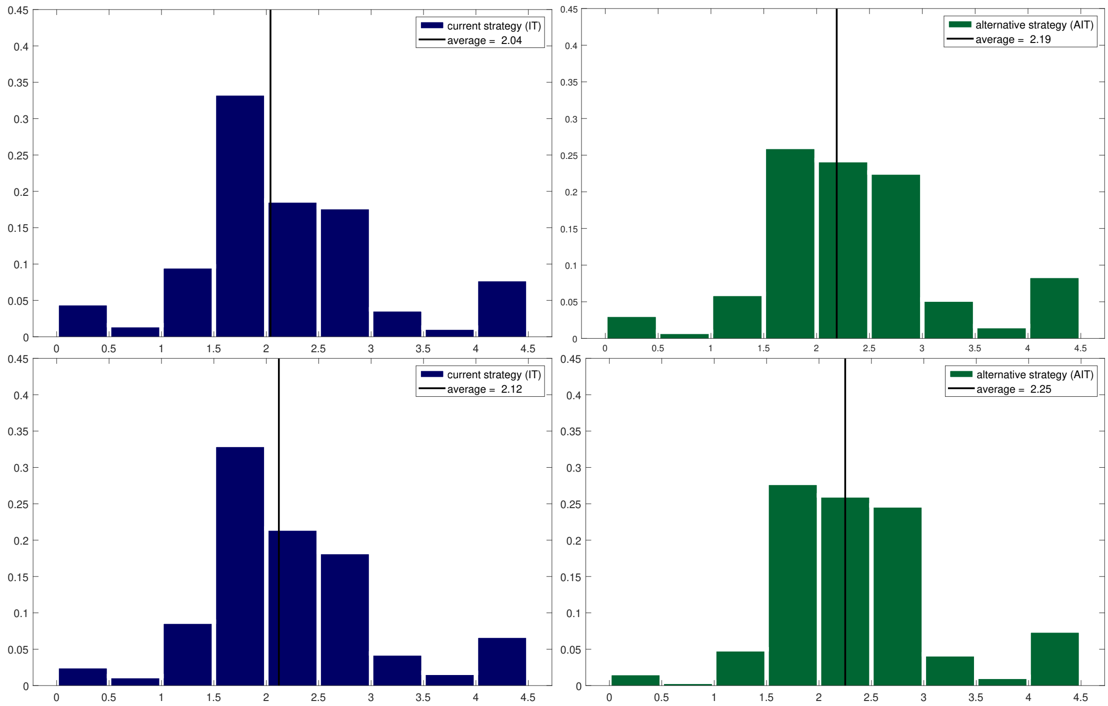


Figure 5.7: Mean inflation expectations 5-10 years ahead (January 2021 Wave 13 and February 2021 Wave 14): IT vs. AIT. Top row: January 2021 (Wave 13) of BOP-HH. Bottom row: February 2021 (Wave 14) of BOP-HH. Left panels: Dark blue bars show mean inflation expectations for the longer term with respondents assuming IT. Right panels: Dark green bars show mean inflation expectations for the longer term with respondents assuming AIT. An upright black line depicts the respective average of the mean inflation expectations. Test statistics are provided in Table 5.5.

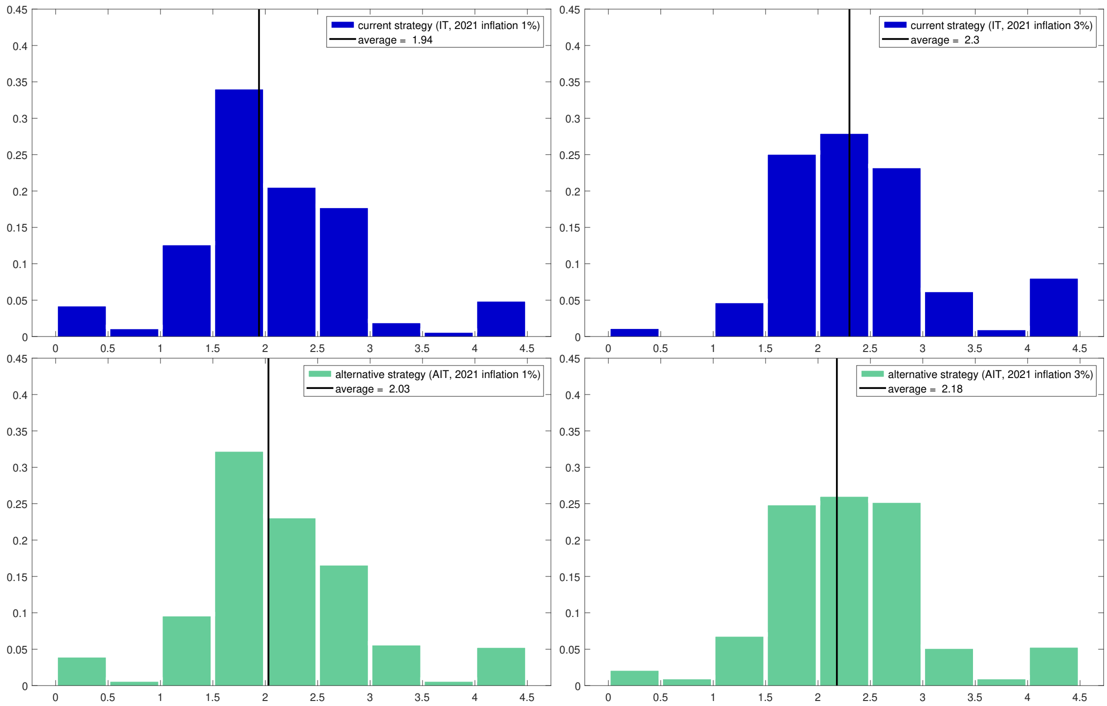


Figure 5.8: Mean Inflation expectations 2-3 years ahead, 2021 inflation at 1% and 3%, respectively (February 2021 Wave 14). February 2021 (Wave 14) of BOP-HH. Left panels: Respondents are asked to assume that 2021 inflation is at 1%. Right panels: Respondents are asked to assume that 2021 inflation is at 3%. All panels: Royal blue bars show mean inflation expectations for the medium term with respondents assuming IT. Teal bars show mean inflation expectations for the medium term with respondents assuming AIT. An upright black line depicts the respective average of the mean inflation expectations. Test statistics are provided in Table 5.5.

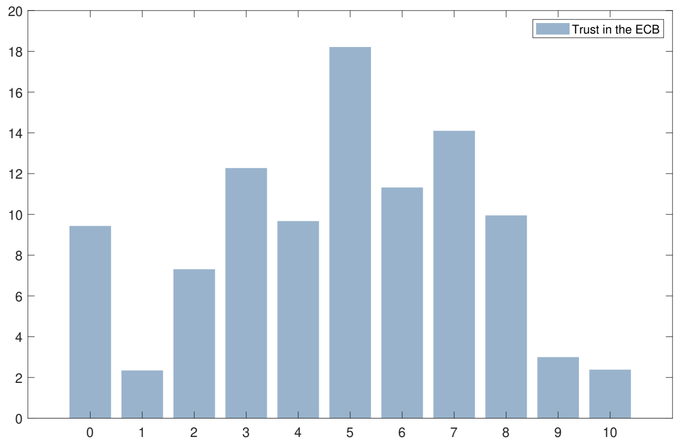


Figure 5.9: To what extent do you trust the ECB's ability to achieve price stability — Distribution of responses. *"On a scale from 0 to 10, how much do you trust that the European Central Bank is able to deliver price stability?* Answer options rank from zero to ten, with 0 = 'Do not trust at all'; 1-9 [gradually increasing trust values]; 10 = 'Trust entirely', or respondents can select 'I don't know the European Central Bank', or 'No answer'.

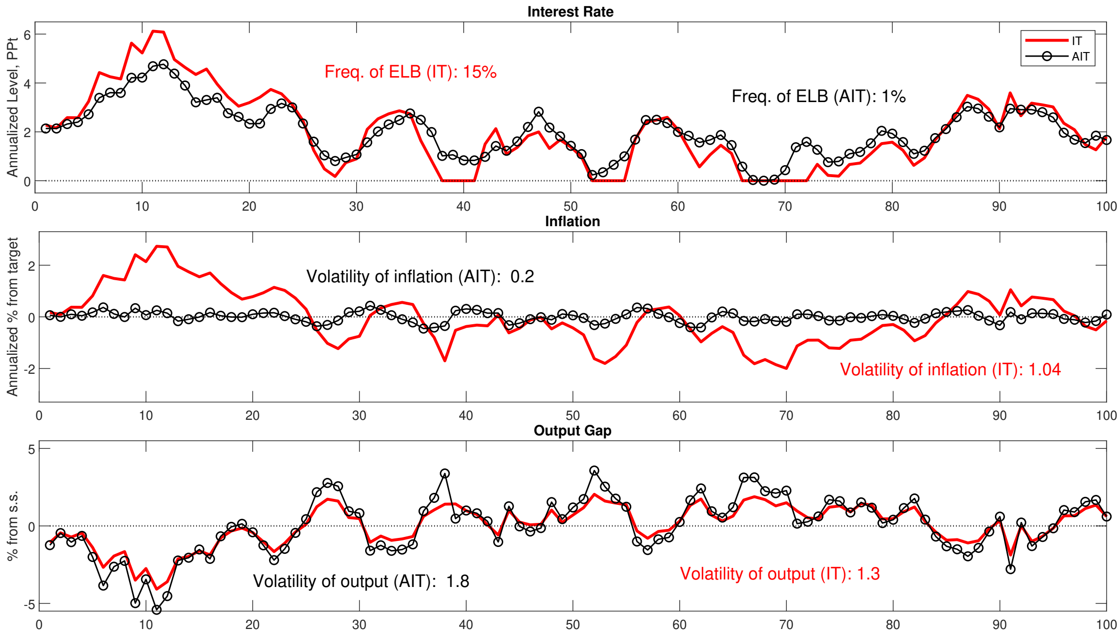


Figure 5.10: Demand shocks dominate supply shocks: Movements of interest rate, inflation, and output. Parameter settings as in Figure 5.4, except that demand shocks dominate supply shocks, $\varepsilon_t^r > \varepsilon_t^u$, by factor 1.5.

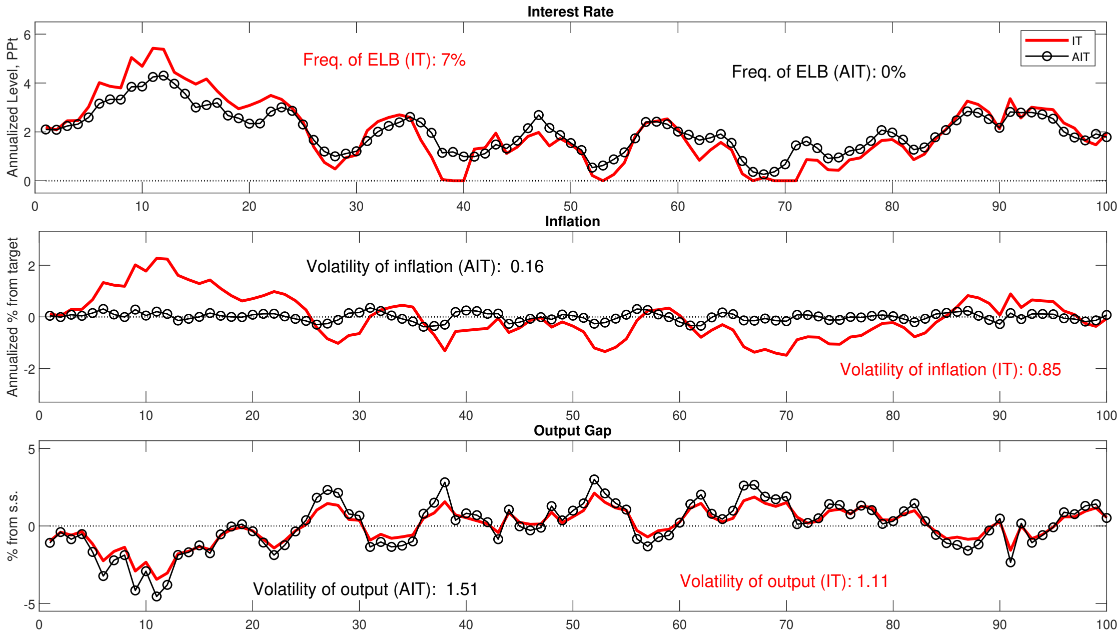


Figure 5.11: Supply shocks dominate demand shocks: Movements of interest rate, inflation, and output. Parameter settings as in Figure 5.4, except that supply shocks dominate demand shocks, $\varepsilon_t^u > \varepsilon_t^r$, by factor 1.5.

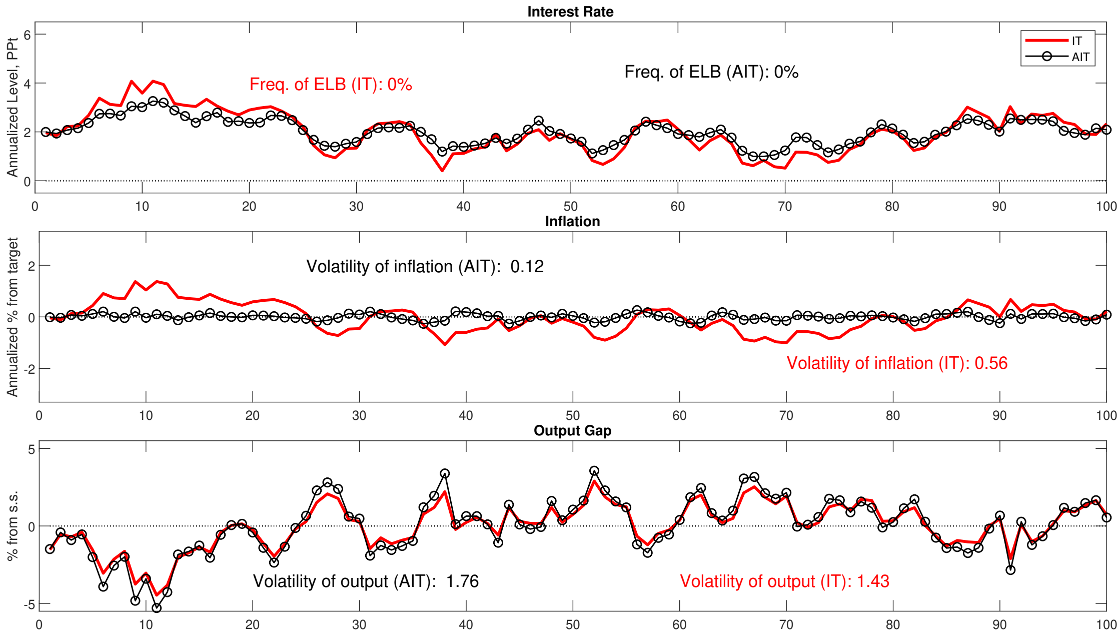


Figure 5.12: The share of credit-constrained households equals $\lambda = 0.4$: Movements of interest rate, inflation, and output. Parameter settings as in Figure 5.4, except that the share of credit-constrained households equals $\lambda = 0.4$.

5.7.2 Additional Tables

Table 5.5: Results for testing equality of mean average subjective probabilities.

Bin considered	$\pi < 1$	$1 \leq \pi < 2$	$2 \leq \pi < 3$	$\pi \geq 3$	jointly
	Wave 10 (medium-term inflation expectations, 2-3 years ahead)				
$IT = IT1\%$	0.335	0.039	0.287	0.040	0.095
$AIT = AIT1\%$	0.008	0.370	0.401	0.030	0.024
$IT = AIT$	0.013	0.008	0.025	0.009	0.002
$IT1\% = AIT1\%$	0.885	0.002	0.059	0.098	0.022
	Wave 13 (longer-term inflation expectations, 5-10 years ahead)				
$IT = IT1\%$	0.699	0.733	0.817	0.890	0.961
$AIT = AIT1\%$	0.274	0.123	0.284	0.178	0.234
$IT = AIT$	0.001	0.000	0.000	0.053	0.000
$IT1\% = AIT1\%$	0.169	0.033	0.002	0.917	0.016
	Wave 14 (medium-term inflation expectations, 2-3 years ahead)				
$IT = IT1\%$	0.043	0.896	0.986	0.071	0.094
$IT = IT3\%$	0.000	0.000	0.000	0.000	0.000
$AIT = AIT1\%$	0.002	0.757	0.032	0.975	0.011
$AIT = AIT3\%$	0.677	0.015	0.146	0.317	0.109
$IT = AIT$	0.003	0.019	0.000	0.560	0.000
$IT1\% = AIT1\%$	0.375	0.098	0.402	0.052	0.158
$IT3\% = AIT3\%$	0.010	0.581	0.973	0.037	0.022
	Wave 14 (longer-term inflation expectations, 5-10 years ahead)				
$IT = IT1\%$	0.509	0.466	0.321	0.504	0.567
$AIT = AIT1\%$	0.047	0.404	0.031	0.770	0.085
$IT = AIT$	0.004	0.000	0.000	0.171	0.000
$IT1\% = AIT1\%$	0.934	0.007	0.042	0.393	0.053

Note: Values in the table correspond to p values for the usual t test statistic when testing for equality of the mean average subjective probabilities from two monetary policy regimes bin by bin. The right column of the table shows p values for the Hotelling (1931) T^2 test statistic when testing for the equality of the mean average subjective probabilities from two monetary policy regimes for all bins jointly.

Table 5.6: Baseline regression results for mean inflation expectations 2-3 years ahead (February 2021 Wave 14)

	$\text{mean}_i^s - \text{mean}_i^{IT} = \sum_s^S \delta_s d_{s,i} + u_i,\ s \in \{AIT, IT1\%, AIT1\%, IT3\%, AIT3\%\}$			
	(1)	(2)	(3)	(4)
AIT	0.12^{***}	0.13^{***}	0.13^{***}	0.10^{***}
	(0.03)	(0.03)	(0.02)	(0.01)
IT1%	-0.08^{***}	-0.09^{***}	-0.09^{***}	-0.05^{**}
	(0.03)	(0.03)	(0.02)	(0.02)
IT3%	0.23^{***}	0.25^{***}	0.25^{***}	0.15^{***}
	(0.03)	(0.03)	(0.04)	(0.02)
AIT1%	0.02	0.03	0.03	0.00
AIT3%	0.22^{***}	0.28^{***}	0.28^{***}	0.16^{***}
	(0.03)	(0.03)	(0.03)	(0.02)
AIT1% − IT1%	0.10^{**}	0.12^{**}	0.12^{***}	0.05^{**}
AIT3% − IT3%	−0.01	0.03	0.03	0.01
Observations	2970	2970	2970	2970
Weighting	No	Yes	Yes	Yes
Robust SE	No	No	Yes	No
Huber weighting	No	No	No	Yes

Note: Sub- and superscript s refers to a certain monetary policy strategy with specific assumptions about current inflation. For instance, $s = AIT3\%$ denotes 'average inflation targeting', with respondents assuming the inflation rate in 2021 at 3%. Asterisks ($^{***}, ^{**}, ^{*}$) indicate significance levels of 1, 5 and 10% for the usual t test statistics for coefficient estimates and for the F test statistics for differences in coefficient estimates.

Table 5.7: Baseline regression results for mean inflation expectations 2-3 years and 5-10 years ahead (October 2020, January 2021, February 2021)

	$\text{mean}_i^s - \text{mean}_i^{IT} = \sum_s^S \delta_s d_{s,i} + u_i,\ s \in \{AIT, IT1\%, AIT1\%, IT3\%, AIT3\%\}$			
	Expectations 2-3 years ahead		Expectations 5-10 years ahead	
	October 2020	February 2021	January 2021	February 2021
	(1)	(2)	(3)	(4)
AIT	0.05^{*}	0.13^{***}	0.15^{***}	0.12^{***}
	(0.03)	(0.03)	(0.03)	(0.03)
IT1%	-0.15^{***}	-0.09^{***}	-0.11^{***}	0.02
	(0.03)	(0.03)	(0.03)	(0.03)
IT3%		0.25^{***}		
		(0.03)		
AIT1%	0.06^{**}	0.03	0.06^{**}	0.07^{**}
	(0.03)	(0.03)	(0.03)	(0.03)
AIT3%		0.28^{***}		
		(0.03)		
AIT1% − IT1%	0.21^{***}	0.12^{**}	0.17^{***}	0.05
AIT3% − IT3%		0.03		
Observations	1903	2970	2342	1745

Note: Sub- and superscript s refers to a certain monetary policy strategy with specific assumptions about current inflation. For instance, $s = AIT3\%$ denotes 'average inflation targeting', with respondents assuming the inflation rate in 2021 at 3%. All regressions use weighted survey responses where the weights ensure the representativeness of the sample. Asterisks ($^{***}, ^{**}, ^{*}$) indicate significance levels of 1, 5 and 10% for the usual t test statistics for coefficient estimates and for the F test statistics for differences in coefficient estimates.

Table 5.8: RCT including follow-up questions on spending intentions of private households

Stage 1	Infobox for all participants: [Infobox as in Table 5.1 of main paper]
Stage 2	All participants — assuming ECB is pursuing current strategy — are asked to assign probabilities for inflation 2-3 years ahead [Histogram as in Table 5.1 of main paper] Participants receive follow-up questions: *"You expect the inflation rate over the next two to three years to [keeps displaying histogram]. Assume that you would like to make major purchases (e.g. a fridge, sofa, or wardrobe). In view of your expectations regarding the inflation rate, which of the following statements applies to you?"* *(1) I think that now would be the right time to make major purchases.* *(2) I think that now would not be the right time to make major purchases.* *"Why do you think that now would be the right time to make major purchases? Please select the reason you think is most important.*
Stage 3	Participants are randomly sampled into one of five subgroups, A, B, C, D$^{\sharp:}$ or E$^{\sharp:}$, facing different assumptions about monetary policy and current inflation. Then, participants are asked again to assign probabilities as in Stage 2: [Histogram as in Table 5.1 of main paper] Participants again receive follow-up questions: *"You expect the inflation rate over the next two to three years to [keeps displaying histogram]. Assume that you would like to make major purchases (e.g. a fridge, sofa, or wardrobe). In view of your expectations regarding the inflation rate, which of the following statements applies to you?"* *(1) I think that now would be the right time to make major purchases.* *(2) I think that now would not be the right time to make major purchases.* *"Why do you think that now would be the right time to make major purchases? Please select the reason you think is most important.*

Note: The RCT with follow-up questions to each subgroup A, B, C, D, and E was conducted exclusively in February 2021 to elicit the spending behavior related to medium-term (2-3 years ahead) expectations. The original questionnaire is in German language.

Table 5.9: Inflation expectations and their role for the readiness to spend on durables.

	$y_i = \alpha + \beta mean_i^s + \gamma_k \mathbf{X_i} + u_i$			
AME	(1)	(2)	(3)	(4)
$Mean_i^s$	0.00	0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.02)
Uncertainty σ_i^s			0.00	
			(0.04)	
Female	0.04^{**}	0.00	0.00	0.06^{**}
	(0.02)	(0.02)	(0.02)	(0.03)
Age				
40 to 60	0.00	0.07^{*}	0.07^{*}	0.00
	(0.03)	(0.04)	(0.04)	(0.05)
over 60	0.00	0.12^{**}	0.12^{**}	0.00
	(0.04)	(0.05)	(0.05)	(0.06)
HH income				
1500 to 3000	0.00	0.03	0.00	0.10
	(0.04)	(0.05)	(0.05)	(0.06)
3000 to 5000	0.10^{**}	0.11^{**}	0.11^{**}	0.17^{**}
	(0.04)	(0.06)	(0.06)	(0.06)
over 5000	0.14^{***}	0.14^{***}	0.14^{**}	0.17^{**}
	(0.04)	(0.06)	(0.06)	(0.07)
HH size				
two	0.00	0.00	0.00	0.00
	(0.03)	(0.03)	(0.03)	(0.04)
three	0.00	0.00	0.00	0.00
	(0.03)	(0.04)	(0.04)	(0.06)
over four	0.00	0.00	0.00	0.09^{*}
	(0.03)	(0.04)	(0.04)	(0.05)
Employed	0.00	0.00	0.00	0.00
	(0.03)	(0.04)	(0.04)	(0.04)
College degree	0.00	0.00	0.00	0.00
	(0.03)	(0.03)	(0.03)	(0.04)
High school degree	0.05^{**}	0.11^{***}	0.11^{***}	0.10^{***}
	(0.03)	(0.03)	(0.03)	(0.04)
East pre 1990	-0.09^{***}	-0.09^{***}	-0.09^{***}	0.00
	(0.02)	(0.03)	(0.03)	(0.04)
Observations	2657	2657	2657	1651
Weighting	No	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes
Adjusters only	No	No	No	Yes

Note: Columns 1 to 4 report the average marginal effects (AME) from a standard probit estimation when the readiness to spend y_i is regressed on the mean inflation expectations post-treatment and a set of controls as shown in the various rows of the table.

Table 5.10: Pre- and post-treatment frequency distribution of the reasons respondents selected to indicate whether or not currently is a good time to buy durable goods.

Pre-treatment	All	*IT1%*	*IT3%*	*AIT*	*AIT1%*	*AIT3%*
Good time to buy = 'Yes'						
Stronger price increase	33.67	35.77	34.61	31.01	34.58	32.29
Stronger price increase in these goods	10.19	10.33	8.91	10.34	10.45	10.94
Need for replacement	23.74	22.67	23.92	26.1	23.13	22.92
Abundant finance	21.19	18.64	21.12	22.22	20.65	23.44
Favourable access to credit	4.58	4.79	4.83	4.65	3.98	4.69
Other	6.62	7.81	6.62	5.68	7.21	5.73
Good time to buy = 'No'						
Weaker price increase	2.12	2.06	1.52	1.57	4.55	0.96
Weaker price increase in these goods	3.24	4.12	3.03	2.09	4.04	2.88
No need for replacement	54.5	59.28	52.02	57.07	50.51	53.85
Not enough finances	13.04	11.34	12.12	12.04	15.15	14.42
No access to credit	5.56	6.19	4.55	5.76	3.03	8.17
Do not spend savings	17.9	14.43	22.73	17.28	18.69	16.35
Other	3.64	2.58	4.04	4.19	4.04	3.37
Post-treatment	All	*IT1%*	*IT3%*	*AIT*	*AIT1%*	*AIT3%*
Good time to buy = 'Yes'						
Stronger price increase	33.57	34.6	32.28	31.94	35.98	32.9
Stronger price increase in these goods	13.89	16.41	13.39	13.61	12.16	13.88
Need for replacement	23.12	21.46	22.57	25.39	24.57	21.59
Abundant finance	20.04	16.41	22.83	19.9	18.36	22.88
Favourable access to credit	4.25	4.04	4.46	5.24	3.47	4.11
Other	5.13	7.07	4.46	3.93	5.46	4.63
Good time to buy = 'No'						
Weaker price increase	2.87	1.47	1.94	2.93	3.63	4.41
Weaker price increase in these goods	3.06	4.9	2.43	2.93	3.63	1.47
No need for replacement	55.93	60.78	55.34	53.66	53.37	56.37
Not enough finances	10.77	7.84	9.71	10.24	13.99	12.25
No access to credit	6.42	4.9	6.31	10.73	3.11	6.86
Do not spend savings	17.29	15.69	20.87	16.59	17.62	15.69
Other	3.66	4.41	3.4	2.93	4.66	2.94

Bibliography

ABEL, J., R. RICH, J. SONG, AND J. TRACY (2016): “The Measurement and Behavior of Uncertainty: Evidence from the ECB Survey of Professional Forecasters,” *Journal of Applied Econometrics*, 31, 533–550.

ALTIG, D., J. M. BARRERO, N. BLOOM, S. J. DAVIS, B. MEYER, AND N. PARKER (2022): “Surveying business uncertainty,” *Journal of Econometrics*, 231, 282–303, Annals Issue: Subjective Expectations & Probabilities in Economics.

AMANO, R., S. GNOCCHI, S. LEDUC, AND J. WAGNER (2020): “Average Is Good Enough: Average Inflation Targeting and the ELB,” Working Paper 2020-21, Federal Reserve Bank of San Francisco.

ANDRADE, P., E. GAUTIER, AND E. MENGUS (2020): “What Matters in Households’ Inflation Expectations?” Working Paper 770, Banque de France.

ANDRADE, P., E. GHYSELS, AND J. IDIER (2012): “Tails of Inflation Forecasts and Tales of Monetary Policy,” Working paper 407, Banque de France.

ANDREWS, D. (1991): “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” *Econometrica*, 59, 817–58.

ARMANTIER, O., W. BRUINE DE BRUIN, S. POTTER, G. TOPA, W. VAN DER KLAAUW, AND B. ZAFAR (2013): “Measuring Inflation Expectations,” *Annual Review of Economics*, 5, 273–301.

ARMANTIER, O., W. B. DE BRUINE, W. VAN DER KLAAUW, G. TOPA, AND B. ZAFAR (2015): “Inflation Expectations and Behavior: Do Survey Respondents Act on their Beliefs?” *International Economic Review*, 56, 505–536.

Armantier, O., G. Koşar, R. Pomerantz, D. Skandalis, K. Smith, G. Topa, and W. van der Klaauw (2021): "How economic crises affect inflation beliefs: Evidence from the Covid-19 pandemic," *Journal of Economic Behavior & Organization*, 189, 443–469.

Armantier, O., G. Topa, W. van der Klaauw, and B. Zafar (2017): "An overview of the Survey of Consumer Expectations," *Federal Reserve Bank of New York Economic Policy Review*, 23, 51–72.

Bachmann, R., T. O. Berg, and E. R. Sims (2015): "Inflation expectations and readiness to spend: Cross-sectional evidence," *American Economic Journal: Economic Policy*, 7, 1–35.

Bachmann, R., K. Carstensen, S. Lautenbacher, and M. Schneider (2021): "Uncertainty and change: survey evidence of firms' subjective beliefs," Working Paper 29430, National Bureau of Economic Research.

Baker, S. R., N. Bloom, and S. J. Davis (2016): "Measuring Economic Policy Uncertainty," *The Quarterly Journal of Economics*, 131, 1593–1636.

Becker, C., P. Duersch, and T. Eife (2023): "Measuring Inflation Expectations: How the Response Scale Shapes Density Forecasts," Working paper, University of Heidelberg, Department of Economics.

Becker, C., P. Duersch, T. A. Eife, and A. Glas (2022): "Extending the procedure of Engelberg et al. (2009) to surveys with varying interval-widths," FAU Discussion Papers in Economics No. 05/2022, Friedrich-Alexander-Universität Erlangen-Nürnberg.

Beckmann, E. and T. Schmidt (2020): "Bundesbank online pilot survey on consumer expectations 01/2020," Tech. rep., Bundesbank Technical Paper.

Ben-David, I., E. Fermand, C. M. Kuhnen, and G. Li (2019): "Expectations uncertainty and household economic behavior," Working Paper 25336, National Bureau of Economic Research.

Ben-David, I., J. R. Graham, and C. R. Harvey (2013): "Managerial miscalibration," *The Quarterly Journal of Economics*, 128, 1547–1584.

Bilbiie, F. O. (2019): "Monetary policy and heterogeneity: An analytical framework," Discussion Paper No. DP12601, CEPR.

Binder, C., T. S. McElroy, and X. S. Sheng (2022): "The Term Structure of Uncertainty: New Evidence from Survey Expectations," *Journal of Money, Credit and Banking*, 54, 39–71.

Binder, C. C. (2017): "Measuring uncertainty based on rounding: New method and application to inflation expectations," *Journal of Monetary Economics*, 90, 1–12.

Blanchflower, D. G. and C. MacCoille (2009): "The formation of inflation expectations: an empirical analysis for the UK," Working Paper 15388, National Bureau of Economic Research.

Blinder, A. S. (2000): "Central-Bank Credibility: Why Do We Care? How Do We Build It?" *American Economic Review*, 90, 1421–1431.

Bloom, N. (2009): "The Impact of Uncertainty Shocks," *Econometrica*, 77, 623–685.

Bobeica, E., M. Ciccarelli, and I. Vansteenkiste (2019): "The link between labor cost and price inflation in the euro area," ECB Working Paper Series 2235.

Boero, G., J. Smith, and K. F. Wallis (2011): "Scoring rules and survey density forecasts," *International Journal of Forecasting*, 27, 379–393.

——— (2015): "The Measurement and Characteristics of Professional Forecasters' Uncertainty," *Journal of Applied Econometrics*, 30, 1029–1046.

Brier, G. W. (1950): "Verification of forecasts expressed in terms of probability," *Monthly Weather Review*, 78, 1–3.

Bruine de Bruin, W., M. Baldassi, B. Figner, B. Fischhoff, L. Fleishman, D. Hardisty, E. Johnson, G. Keren, M. Konnikova, I. Levin, et al. (2011): "Framing effects in surveys: How respondents make sense of the questions we ask," in *Perspectives on Framing, edited by Gideon Keren*, New-York, NY. Taylor and Francis Group, Psychology Press, 303–325.

Bruine de Bruin, W., B. Fischhoff, S. G. Millstein, and B. L. Halpern-Felsher (2000): "Verbal and numerical expressions of probability:"It's a fifty–fifty chance"," *Organizational behavior and human decision processes*, 81, 115–131.

Bruine de Bruin, W., W. van der Klaauw, J. S. Downs, B. Fischhoff, G. Topa, and O. Armantier (2010): "Expectations of inflation: The role of demographic variables, expectation formation, and financial literacy," *Journal of Consumer Affairs*, 44, 381–402.

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

Bruine de Bruin, W., W. van der Klaauw, M. van Rooij, F. Teppa, and K. de Vos (2017): "Measuring expectations of inflation: Effects of survey mode, wording, and opportunities to revise," *Journal of Economic Psychology*, 59, 45–58.

Bryan, M. F. and G. Venkatu (2001): "The curiously different inflation perspectives of men and women," *Economic Commentary*.

Candia, B., O. Coibion, and Y. Gorodnichenko (2020): "Communication and the beliefs of economic agents," Working Paper 27800, National Bureau of Economic Research.

Carriero, A., T. E. Clark, and M. Marcellino (2018): "Measuring uncertainty and its impact on the economy," *The Review of Economics and Statistics*, 100, 799–815.

Carvalho, C., S. Eusepi, E. Moench, and B. Preston (2023): "Anchored inflation expectations," *American Economic Journal: Macroeconomics*, 15, 1–47.

Cascaldi-Garcia, D., D. D. Datta, T. R. T. Ferreira, O. V. Grishchenko, M. R. Jahan-Parvar, J. M. Londono, F. Loria, S. Ma, M. del Giudice Rodriguez, J. H. Rogers, C. Sarisoy, and I. Zer (2021): "What is Certain about Uncertainty?" Forthcoming in the Journal of Econometric Literature.

Cavallo, A., G. Cruces, and R. Perez-Truglia (2017): "Inflation expectations, learning, and supermarket prices: Evidence from survey experiments," *American Economic Journal: Macroeconomics*, 9, 1–35.

Christelis, D., D. Georgarakos, T. Jappelli, and M. van Rooij (2020): "Trust in the Central Bank and Inflation Expectations," *International Journal of Central Banking*, 16, 1–37.

Clarida, R. (2019): "The Federal Reserve's Review of Its Monetary Policy Strategy, Tools, and Communication Practices," A Hot Economy: Sustainability and Trade-Offs, San Francisco Federal Reserve conference 26 September.

Clark, T., G. Ganics, and E. Mertens (2022): "What is the Predictive Value of SPF Point and Density Forecasts?" Working Paper 22-37, Federal Reserve Bank of Cleveland.

Clark, T. E. and F. Ravazzolo (2015): "Macroeconomic Forecasting Performance under Alternative Specifications of Time-Varying Volatility," *Journal of Applied Econometrics*, 30, 551–575.

Clements, M. P. (2014): "Forecast Uncertainty—Ex Ante and Ex Post: U.S. Inflation and Output Growth," *Journal of Business & Economic Statistics*, 32, 206–216.

Clements, M. P., R. W. Rich, and J. S. Tracy (2023): "Surveys of professionals," in *Handbook of Economic Expectations*, ed. by R. Bachmann, G. Topa, and W. van der Klaauw, Academic Press, chap. 3, 71–106.

Coenen, G. and R. Straub (2005): "Does Government Spending Crowd in Private Consumption? Theory and Empirical Evidence for the Euro Area," *International Finance*, 8, 435–470.

Coibion, O., D. Georgarakos, Y. Gorodnichenko, G. Kenny, and M. Weber (2022a): "The effect of macroeconomic uncertainty on household spending," Chicago Booth Research Paper No. 22-17, University of Chicago, Becker Friedman Institute for Economics.

Coibion, O., D. Georgarakos, Y. Gorodnichenko, and M. Van Rooij (2019): "How Does Consumption Respond To News About Inflation? Field Evidence From A Randomized Control Trial," Working Paper 26106, National Bureau of Economic Research.

Coibion, O., Y. Gorodnichenko, E. S. Knotek II, and R. Schoenle (2021): "Average Inflation Targeting And Household Expectations," Working Paper 27836, National Bureau of Economic Research.

Coibion, O., Y. Gorodnichenko, and S. Kumar (2018a): "How do firms form their expectations? New survey evidence," *American Economic Review*, 108, 2671–2713.

Coibion, O., Y. Gorodnichenko, S. Kumar, and M. Pedemonte (2020): "Inflation expectations as a policy tool?" *Journal of International Economics*, 124, 103297.

Coibion, O., Y. Gorodnichenko, S. Kumar, and J. Ryngaert (2018b): "Do You Know That I Know That You Know ...? Higher-Order Beliefs in Survey Data," Working Paper 24987, National Bureau of Economic Research.

Coibion, O., Y. Gorodnichenko, and M. Weber (2022b): "Monetary policy communications and their effects on household inflation expectations," *Journal of Political Economy*, 130, 1537–1584.

Croushore, D. and T. Stark (2019): "Fifty years of the survey of professional forecasters," *Economic Insights*, 4, 1–11.

Croushore, D. D. (1993): "Introducing: The Survey of Professional Forecasters," *Federal Reserve Bank of Philadelphia Business Review*, 6, 3–15.

Curtin, R. (1996): "Procedure to estimate price expectations," Tech. rep., Ann Arbour, MI: University of Michigan Survey Research Center.

D'Acunto, F., D. Hoang, M. Paloviita, and M. Weber (2020): "Effective policy communication: Targets versus instruments," KIT Working Paper Series in Economics No. 147.

D'Acunto, F., U. Malmendier, and M. Weber (2023): "What do the data tell us about inflation expectations?" in *Handbook of economic expectations*, Elsevier, 133–161.

Deutsche Bundesbank (2022): "Survey on consumer expectations," `https://www.bundesbank.de/en/bundesbank/research/survey-on-consumer-expectations/survey-on-consumer-expectations-794568` (last accessed: March 1, 2023).

Dominitz, J. and C. F. Manski (1997): "Using expectations data to study subjective income expectations," *Journal of the American Statistical Association*, 92, 855–867.

Dräger, L. and U. Fritsche (2013): "Don't Worry, Be Right! Survey Wording Effects on In flation Perceptions and Expectations," DEP (Socioeconomics) Discussion Papers-Macroeconomics and Finance Series No.8/ 2013.

D'Acunto, F., U. Malmendier, J. Ospina, and M. Weber (2021a): "Exposure to grocery prices and inflation expectations," *Journal of Political Economy*, 129, 1615–1639.

D'Acunto, F., U. Malmendier, and M. Weber (2021b): "Gender roles produce divergent economic expectations," Proceedings of the National Academy of Sciences 21.

ECB (2019): "Inflation expectations and the conduct of monetary policy," Speech by Benoît Cœuré, Member of the Executive Board of the ECB, at an event organised by the SAFE Policy Center, Frankfurt am Main, 11 July 2019. Available at `https://www.ecb.europa.eu/press/key/date/2019/html/ecb.sp190711~6dcaf97c01.en.html` (last accessed: August 8, 2019).

Ehm, W., T. Gneiting, A. Jordan, and F. Krüger (2016): "Of quantiles and expectiles: consistent scoring functions, Choquet representations and forecast rankings," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78, 505–562.

Engelberg, J., C. F. Manski, and J. Williams (2009): "Comparing the Point Predictions and Subjective Probability Distributions of Professional Forecasters," *Journal of Business & Economic Statistics*, 27, 30–41.

Epstein, E. S. (1969): "A scoring system for probability forecasts of ranked categories," *Journal of Applied Meteorology*, 8, 985–987.

Falck, E., M. Hoffmann, and P. Huertgen (2021): "Disagreement about Inflation Expectations and Monetary Policy Transmission," *Journal of Monetary Economics*, 118, 15–31.

Federal Reserve Bank of New York (2020): "Survey of Consumer Expectations," `https://www.newyorkfed.org/microeconomics/sce` (last accessed: March 9, 2020).

Federal Reserve Bank of Philadelphia (2022): "Survey of Professional Forecasters," `https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/survey-of-professional-forecasters` (last accessed: July 19, 2022).

Fischhoff, B. and W. Bruine de Bruin (1999): "Fifty–fifty= 50%?" *Journal of Behavioral Decision Making*, 12, 149–163.

Friedman, M. (1937): "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance," *Journal of the American Statistical Association*, 32, 675–701.

Galí, J. and M. Gertler (1999): "Inflation dynamics: A structural econometric analysis," *Journal of Monetary Economics*, 44, 195–222.

Galí, J., J. López-Salido, and J. Vallés (2007): "Understanding the Effects of Government Spending on Consumption," *Journal of the European Economic Association*, 5, 227–270.

Galvao, A. B. and J. Mitchell (2019): "Measuring data uncertainty: An application using the Bank of England's "fan charts" for historical GDP growth," Discussion Paper 2019-08, ESCoE.

Ganics, G., B. Rossi, and T. Sekhposyan (2020): "From Fixed-Event to Fixed-Horizon Density Forecasts: Obtaining Measures of Multi-Horizon Uncertainty from Survey Density Forecasts," Working Paper 1142, Barcelona School of Economics, forthcoming in the Journal of Money, Credit and Banking.

Garcia, J. A. (2003): "An introduction to the ECB's Survey of Professional Forecasters," Occasional Paper Series 8, European Central Bank.

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

Glas, A. (2020): "Five dimensions of the uncertainty–disagreement linkage," *International Journal of Forecasting*, 36, 607–627.

Glas, A. and M. Hartmann (2022): "Uncertainty measures from partially rounded probabilistic forecast surveys," *Quantitative Economics*, 13, 979–1022.

Gneiting, T., F. Balabdaoui, and A. E. Raftery (2007): "Probabilistic forecasts, calibration and sharpness," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69, 243–268.

Gneiting, T. and M. Katzfuss (2014): "Probabilistic forecasting," *Annual Review of Statistics and Its Application*, 1, 125–151.

Gneiting, T. and A. E. Raftery (2007): "Strictly proper scoring rules, prediction, and estimation," *Journal of the American Statistical Association*, 102, 359–378.

Gneiting, T. and R. Ranjan (2013): "Combining predictive distributions," *Electronic Journal of Statistics*, 7, 1747–1782.

Gneiting, T. and J. Resin (2022): "Regression diagnostics meets forecast evaluation: Conditional calibration, reliability diagrams, and coefficient of determination," Working paper, available at `https://arxiv.org/abs/2108.03210v3` (last accessed: March 2, 2023).

Goldfayn-Frank, O. and J. Wohlfart (2020): "Expectation formation in a new environment: Evidence from the German reunification," *Journal of Monetary Economics*, 115, 301–320.

Good, I. (1952): "Rational decisions," *Journal of the Royal Statistical Society. Series B (Methodological)*, 14, 107–114.

Gosselin, M.-A. and M. Khan (2015): "A Survey of Consumer Expectations for Canada," *Bank of Canada Review*, 2015, 14–23.

Grishchenko, O., S. Mouabbi, and J.-P. Renne (2019): "Measuring Inflation Anchoring and Uncertainty: A US and Euro Area Comparison," *Journal of Money, Credit and Banking*, 51, 1053–1096.

Guerrieri, L. and M. Iacoviello (2015): "OccBin: A toolkit for solving dynamic models with occasionally binding constraints easily," *Journal of Monetary Economics*, 70, 22–38.

Gómez, V. and A. Maravall (1998): "Guide for Using the Programs TRAMO and SEATS (Beta Version: December 1997)," Working Paper 9805, Banco de España.

Hammond, G. et al. (2012): "State of the art of inflation targeting," Handbooks 29, Centre for Central Banking Studies, Bank of England.

Hayo, B. and P.-G. Méon (2022): "Measuring household inflation perceptions and expectations: The effect of guided vs non-guided inflation questions," Working Paper No. 11-2022, MAGKS Joint Discussion Paper Series in Economics.

Hoffmann, M., M. Krause, M. Kliem, and R. Sauer (2021a): "Rebalancing the Euro area: Is Wage Adjustment in Germany the Answer?" *Journal of Internatinal Money and Finance*, 119, 102497.

Hoffmann, M., E. Moench, L. Pavlova, and G. Schultefrankenfeld (2021b): "The effects of the ECB's new inflation target on private households' inflation expectations," Research Brief 43, Deutsche Bundesbank.

Hotelling, H. (1931): "The generalization of Student's ratio," *Annals of Mathematical Statistics*, 2, 360–378.

Huber, P. J. (1981): *Robust Statistics*, Wiley, New York.

Huisman, R., N. L. Van der Sar, and R. C. Zwinkels (2021): "Volatility expectations and disagreement," *Journal of Economic Behavior & Organization*, 188, 379–393.

Hurd, M. D. (2009): “Subjective probabilities in household surveys,” *Annual Review of Economics*, 1, 543–562.

Jurado, K., S. C. Ludvigson, and S. Ng (2015): “Measuring Uncertainty,” *American Economic Review*, 105, 1177–1216.

Katona, G. and R. Likert (1946): “Relationship between consumer expenditures and savings: The contribution of survey research,” *The Review of Economics and Statistics*, 28, 197–199.

Knüppel, M. and F. Krüger (2022): “Forecast uncertainty, disagreement, and the linear pool,” *Journal of Applied Econometrics*, 37, 23–41.

Knüppel, M. and A. L. Vladu (2016): “Approximating fixed-horizon forecasts using fixed-event forecasts,” Discussion Papers 28/2016, Deutsche Bundesbank.

Krüger, F. and L. Pavlova (2021): “Quantifying subjective uncertainty in survey expectations,” KIT Working Paper Series in Economics No. 664.

Lahiri, K. and X. Sheng (2010): “Measuring forecast uncertainty by disagreement: The missing link,” *Journal of Applied Econometrics*, 25, 514–538.

López-Menéndez, A. J. and R. Pérez-Suárez (2019): “Acknowledging uncertainty in economic forecasting. Some insight from confidence and industrial trend surveys,” *Entropy*, 21(4), 413.

Ludvigson, S. C., S. Ma, and S. Ng (2021): “Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response?” *American Economic Journal: Macroeconomics*, 13, 369–410.

Mackowiak, B. and M. Wiederholt (2009): “Optimal sticky prices under rational inattention,” *American Economic Review*, 99, 769–803.

Malmendier, U. and S. Nagel (2016): “Learning from inflation experiences,” *The Quarterly Journal of Economics*, 131, 53–87.

Manksi, C. F. (2023): “Looking ahead to research enhancing measurement of expectations,” in *Handbook of economic expectations*, Elsevier, 809–822.

MANSKI, C. F. (2004): "Measuring expectations," *Econometrica*, 72, 1329–1376.

——— (2018): "Survey measurement of probabilistic macroeconomic expectations: progress and promise," *National Bureau of Economic Research Macroeconomics Annual*, 32, 411–471.

MANSKI, C. F. AND F. MOLINARI (2010): "Rounding probabilistic expectations in surveys," *Journal of Business & Economic Statistics*, 28, 219–231.

MATHESON, J. E. AND R. L. WINKLER (1976): "Scoring rules for continuous probability distributions," *Management Science*, 22, 1087–1096.

MAUERSBERGER, F. AND R. NAGEL (2018): "Levels of Reasoning in Keynesian Beauty Contest Models: A Generative Framework," in *Handbook of Computational Economics*, ed. by C. Hommes and B. LeBaron, vol. 4, chap. 10.

MCCONNELL, M. M. AND G. PEREZ-QUIROS (2000): "Output Fluctuations in the United States: What Has Changed Since the Early 1980's?" *The American Economic Review*, 90, 1464–1476.

MELOSI, L. (2017): "Signaling Effects of Monetary Policy," *Review of Economic Studies*, 84, 853–884.

MERTENS, T. AND J. WILLIAMS (2019): "Tying down the anchor: Monetary policy rules and the lower bound on interest rates," Working Paper Series 14, Federal Reserve Bank of San Francisco.

MUTH, J. F. (1961): "Rational expectations and the theory of price movements," *Econometrica: Journal of the Econometric Society*, 315–335.

NAGEL, J. (2022): "Return of inflation: What do we learn from survey data?" Speech at the Joint Spring Conference on Monetary Policy and Expectations of Households and Firms (last accessed: October 14, 2022).

NEWEY, W. K. AND K. D. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Ollech, D. and K. Webel (2023): "A Random Forest-based Approach to Combining and Ranking Seasonality Tests," *Journal of Econometric Methods*, 12, 117–130.

Patton, A. and A. Timmermann (2011): "Predictability of Output Growth and Inflation: A Multi-Horizon Survey Approach," *Journal of Business & Economic Statistics*, 29, 397–410.

Phillot, M. and R. Rosenblatt-Wisch (2018): "Inflation expectations: The effect of question ordering on forecast inconsistencies," SNB Working Paper Series 11/2018, Swiss National Bank.

Ratto, M., W. Roeger, and J. in't Veld (2009): "QUEST III: An estimated Open-Economy DSGE model of the Euro area with Fiscal and Monetary Policy," *Economic Modelling*, 26, 222–233.

Reifschneider, D. and J. Williams (2000): "Three Lessons for Monetary Policy in a Low-Inflation Era," *Journal of Money, Credit and Banking*, 32, 936–966.

Rich, R. and J. Tracy (2010): "The Relationships among Expected Inflation, Disagreement, and Uncertainty: Evidence from Matched Point and Density Forecasts," *The Review of Economics and Statistics*, 92, 200–207.

Rossi, B. and T. Sekhposyan (2015): "Macroeconomic Uncertainty Indices Based on Nowcast and Forecast Error Distributions," *American Economic Review*, 105, 650–55.

Schotter, A. and I. Trevino (2014): "Belief elicitation in the laboratory," *Annual Review of Economics*, 6, 103–128.

Shannon, C. E. (1948): "A mathematical theory of communication," *Bell System Technical Journal*, 27, 379–423.

Sims, C. A. (2003): "Implications of rational inattention," *Journal of Monetary Economics*, 50, 665–690.

Strähl, C. and Ziegel (2017): "Cross-calibration of probabilistic forecasts," *Electronic Journal of Statistics*, 11, 608–639.

Tsyplakov, A. (2013): "Evaluation of probabilistic forecasts: Proper scoring rules and moments," Working paper, available at `https://dx.doi.org/10.2139/ssrn.2236605` (last accessed: June 23, 2022).

Wallis, K. F. (2005): "Combining Density and Interval Forecasts: A Modest Proposal*," *Oxford Bulletin of Economics and Statistics*, 67, 983–994.

Weber, M., F. D'Acunto, Y. Gorodnichenko, and O. Coibion (2022): "The subjective inflation expectations of households and firms: Measurement, determinants, and implications," *Journal of Economic Perspectives*, 36, 157–184.

Wei, W., F. Balabdaoui, and L. Held (2017): "Calibration tests for multivariate Gaussian forecasts," *Journal of Multivariate Analysis*, 154, 216–233.

Zarnowitz, V. and L. A. Lambros (1987): "Consensus and Uncertainty in Economic Prediction," *Journal of Political Economy*, 95, 591–621.

Zhao, Y. (2022): "Internal consistency of household inflation expectations: Point forecasts vs. density forecasts," *International Journal of Forecasting*, `https://doi.org/10.1016/j.ijforecast.2022.08.008`.