// NO.26-017 | 04/2026

DISCUSSION PAPER

// MALTE KNÜPPEL AND LORA PAVLOVA

Survey Design and Professional Forecasters: The Case of Uncertainty in the US SPF

Leibniz Association

ZEW

Survey design and professional forecasters: The case of uncertainty in the US SPF*

Malte Knüppel
Deutsche Bundesbank

Lora Pavlova
ZEW – Leibniz Centre for European Economic Research

April 20, 2026

Abstract

Histogram forecasts of inflation and growth from the US Survey of Professional Forecasters (SPF) allow for an assessment of the evolution of forecast uncertainty. However, this assessment is complicated by structural breaks in measured uncertainty arising from changes in histogram bin widths over time. The existing literature typically does not take these breaks into account. We propose a break adjustment based on the insights provided by a structural break in 2014, during which bin widths—and consequently, measured inflation uncertainty—shifted significantly, despite true inflation uncertainty remaining virtually constant. Drawing on our results, we propose horizon-specific bin widths for inflation and growth to align measured uncertainty more closely with underlying uncertainty.

Keywords: survey forecasts, volatility, structural breaks

JEL classification: C53, E37.

*E-Mail addresses: malte.knueppel@bundesbank.de, lora.pavlova@zew.de. The views expressed in this work are those of the authors and do not necessarily coincide with the views of the Deutsche Bundesbank or the Eurosystem. We are grateful to Alexander Glas, Kajal Lahiri, Xuguang Simon Sheng, and Robert Rich for their valuable comments and suggestions, which greatly improved this work. Lora Pavlova gratefully acknowledges financial support from the German Science Foundation (DFG) via grant KR 5214/1-1.

1 Introduction

The insights of Bloom (2009), together with 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 measurement. Therefore, many uncertainty measures 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 (Binder, 2017), econometric-based (e.g. Jurado et al., 2015), and market-based (e.g., the CBOE Volatility Index, VIX) measures.

In our study, we investigate the survey-based uncertainty measure of the US Survey of Professional Forecasters (US SPF). This survey asks respondents to assign probabilities to mutually exclusive future events, resulting in histogram forecasts. Since the forecast targets are continuous variables, these events are given by their realizations in certain intervals. Many other surveys, like the SPF of the European Central Bank (ECB), the New York Fed's Survey of Consumer Expectations, the Deutsche Bundesbank Online Panel - Households, and the ECB's Consumer Expectations Survey adopt the same approach.

When histogram forecasts are available, measures of average uncertainty can be constructed in two different ways. As pointed out by Clements (2014), one can first calculate an average histogram across respondents and then gauge this histogram's dispersion. Alternatively, 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*. Like most of the literature, we will investigate *aggregate uncertainty*, but our insights are relevant for the uncertainty of the aggregate histogram forecast as well, because aggregate uncertainty is one of its two drivers.[1]

Our study focuses on how breaks in the survey design hamper the analysis of aggregate uncertainty. It turns out that changes in the bin widths, which occurred several times since the start of the SPF, have quantitatively important, permanent effects on measured uncertainty. We propose a break adjustment designed for changes in bin widths to overcome this problem. In addition, we use these insights to propose bin widths that would lead to levels of aggregate uncertainty that, on average, are approximately consistent with the true underlying uncertainty. These bin widths depend on the respective forecast horizon.

Few studies have investigated the implications of changes in the bin definitions of histogram forecasts. In a study on the relation between rounding behavior of survey participants and measured uncertainty, Glas and Hartmann (2022) note that, after the halving of the US SPF's bin widths for inflation in 2014, the number of bins with non-zero probabilities assigned by the respondents increased from 4 to 5 bins on average, instead of an increase from 4 to 8 bins. The halving of the bin widths thus led to a strong decrease in measured un-

[1] 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.

certainty. However, to the best of our knowledge, this fact has hardly been discussed in the literature. Virtually all studies implicitly assume that professional forecasters have a forecast probability distribution in mind and that they simply map this distribution to the histogram bins. However, since the bin width has a strong effect on measured uncertainty, the conclusions of such US SPF-based studies might often be spurious. For instance, discussions of over- or underconfidence of forecasters, as done in Lichtendahl et al. (2013), Clements (2014), or Broer and Kohlhas (2024, Table III) rely on measured uncertainty. Any analysis involving the forecast accuracy of the histogram forecasts as found in Ganics et al. (2024), Diebold et al. (1999), Clements (2018), or Clark et al. (2022) is likewise influenced by the bin widths used in the samples under study. Finally, the results of analyses of uncertainty like Giordani and Söderlind (2003), Lahiri and Sheng (2010), and Rich and Tracy (2010) also depend on the bin widths in the US SPF data employed. In contrast to studies relying on professional forecasters, studies based on consumer surveys have already started to investigate the relation between bin definitions and measured uncertainty (see Becker et al. 2023) and to propose alternatives to histogram forecasts (see Goldfayn-Frank et al. 2025). An alternative in the context of firm surveys can be found in Altig et al. (2022).

Only a few studies of the US SPF attempt to provide results that are robust concerning changes in the bin widths. For instance, Engelberg et al. (2009) focus on a subsample without changes in the survey design. This approach would now be more difficult to justify because such subsamples have become very short compared to the full sample. Liu and Sheng (2019) regress the panel of individual uncertainty measures of US SPF participants on, inter alia, dummies controlling for different survey structures and use the residual as an uncertainty measure. This approach would be valid if the true uncertainty is constant across the subsamples corresponding to the different survey structures. Yet, this assumption is unlikely to have been satisfied at least since the COVID-19 pandemic.[2] Rich and Tracy (2010) use the widest bins for 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. This approach would be valid if each respondent always tried to map her subjective probability distribution to the histogram as precisely as possible, irrespective of the bin width, such that wider bins only imply a loss of information.[3] Moreover, this approach cannot be applied to GDP growth anymore because of the extreme changes in bin widths during the COVID-19 pandemic.

Section 2 describes the US SPF data, the uncertainty measures, and the structural breaks. Break-adjusted uncertainty measures are constructed in Section 3. Section 4 proposes horizon-specific bin widths aligning measured uncertainty with the true underlying uncertainty. Section 5 concludes.

[2]Due to Liu and Sheng's (2019) measure being a residual, its level does not contain information about the level of specific measures of uncertainty expected by a participant, like, for instance, her expected squared forecast error. Therefore, the level of this ex ante measure of uncertainty cannot be compared to any other measures, like, for instance, the calibration uncertainty employed in Section 4.

[3]Conflating narrow bins to obtain wide bins will likely affect the measured uncertainty. Yet, Lahiri and Wang (2020) only find minor effects of this kind for the US SPF.

Table 1: Sample dates and numbers of forecasters

variable	begin sample	end sample	min	average	max
		current year			
GDP deflator growth	1981q3	2025q1	7	32	51
real GDP growth	1981q3	2025q1	7	33	52
core CPI inflation	2007q1	2025q1	24	34	43
core PCE inflation	2007q1	2025q1	24	32	40
		next year			
GDP deflator growth	1981q3	2025q1	7	32	50
real GDP growth	1981q3	2025q1	7	32	50
core CPI inflation	2007q1	2025q1	24	33	43
core PCE inflation	2007q1	2025q1	24	31	39

Note: *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. *current year* and *next year* refer to the forecast horizon.

2 US SPF uncertainty and changes in survey design

2.1 Determining US SPF 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 (see also Croushore and Stark, 2019, for a more detailed overview). The panellists, 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, panellists 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. In 1985q1 and 1986q1, the forecast horizons are not clear, so we exclude those surveys from our dataset.[5]

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

[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

[5] See statement on p.37 in https://www.philadelphiafed.org/-/media/frbp/assets/surveys-and-data/survey-of-professional-forecasters/spf-documentation.pdf

include 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, a distribution-free approach to measure uncertainty based on entropy is proposed by Rich and Tracy (2010) and Krüger and Pavlova (2024).

The variance will be our preferred measure for uncertainty, because the true value of this uncertainty can be unbiasedly estimated ex post using the squared error of the mean forecast. To compute the variance 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. if 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 A.

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 the mean μ and the variance σ^2 such that the average squared difference between the empirical probability mass function (PMF) and the corresponding PMF implied by the normal distribution is minimized, see Appendix A for details. We use different sets of starting values for each optimization and, among the solutions, select the resulting set of μ and σ^2 yielding the lowest average squared difference. 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 twice 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 2025q1, and the sample for core CPI and core PCE inflation covers the period 2007q1 to 2025q1.[6] Table 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

[6] We do not exclude forecasters who only submit a subset of the histograms. For instance, if a forecaster only provides a histogram forecast for current-year real GDP growth, that histogram is included in our analysis. Thus, from all available histograms, only the two-bin histograms with positive probabilities in non-adjacent bins are excluded from the sample, as mentioned above.

change in the level of the GDP deflator (for GDP deflator growth) and of real GDP (for real GDP growth). 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. Hence, the forecasts for these growth rates are fixed-event forecasts as well. In all cases, 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.

Denoting the uncertainty forecast for variable z made in year t and quarter i for the year $t+j$ by forecaster n by $\sigma^z_{n;j|t,i}$, the aggregate uncertainty in that quarter equals

$$\sigma^z_{j|t,i} = \sqrt{\frac{1}{N_{t,i}} \sum_{n=1}^{N_{t,i}} \left(\sigma^z_{n;j|t,i}\right)^2} \tag{1}$$

with $j = 0, 1$ denoting the forecast horizon in years (i.e. 0 for the current year, 1 for the next year) and $N_{t,i}$ denoting the number of forecasters in year t and quarter i.[7] We employ the square root of the average variance rather than the average variance itself, because large values of the average variance in periods like 2020q2, when the COVID-19 pandemic started, can otherwise dominate the scale of figures and obscure variations within the typical range.

The resulting time series of aggregate uncertainty for GDP deflator growth and for real GDP growth are shown in Figure 1. One can see four large changes in uncertainty. For GDP deflator growth, uncertainty dropped markedly in 1992q1 and 2014q1. For real GDP growth, 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.

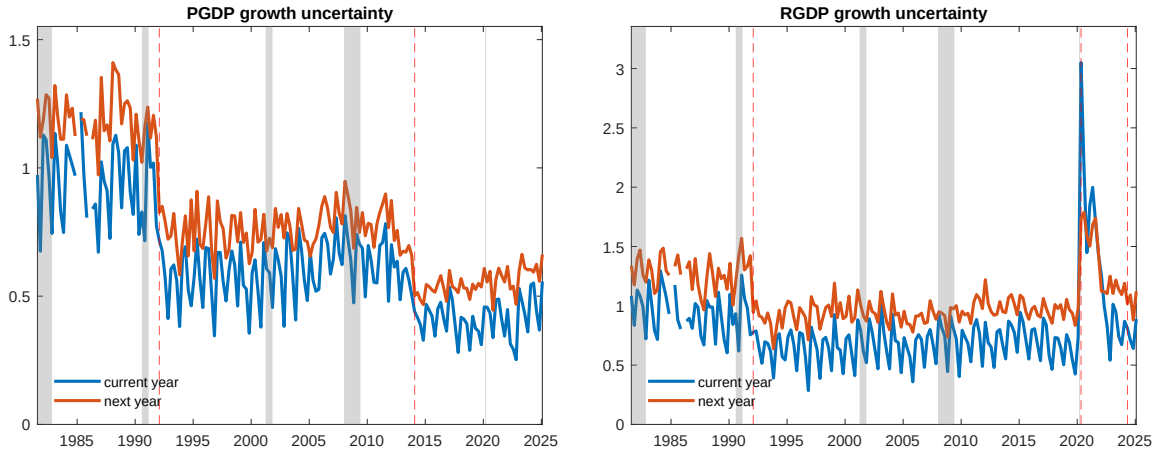

Figure 1: Aggregate uncertainty of GDP deflator growth (left panel) and real GDP growth (right panel) concerning the current and the next year. Uncertainty is measured as defined in (1). The red vertical lines indicate dates with changes in bin widths. Shaded areas indicate recessions as dated by the NBER.

[7] There are rare cases where forecasters only produce histograms for one forecast horizon, such that N also depends on j. For notational convenience, we neglect these cases here.

Figure 2: Current-year GDP deflator (left panel) and real GDP (right panel) growth forecasts (averages of means of individual histograms) together with histogram bins over time. Black dashed lines mark periods with changes in the bin definitions. Bin definitions of US SPF are identical for current- and next-year forecasts. Graph is based on presentation slides for study of Clark et al. (2022).

In 1992q1, the bin width for GDP deflator growth and real GDP growth declined from 2 percentage points (pp) to 1 pp. In 2014q1, the bin width for GDP deflator growth halved again, reaching 0.5 pp. In 2020q2, the bin widths for real GDP growth became larger and heterogenous. 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, reaching 6 pp for the widest of the closed bins. In 2024q2, the width of most bins was halved, but the width of the three central bins remained unchanged. Figure 2 shows how the bin widths and the number of bins changed over time.

The halving of bin widths 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 real GDP 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. Their decrease in 2024q2 was likely due to the confidence that growth volatility had attained lower levels by then. The second halving of bin widths for GDP deflator growth in 2014q1 was probably again due to the low volatility of this variable in previous years.

2.2 The break in measured GDP deflator growth uncertainty in 2014

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.

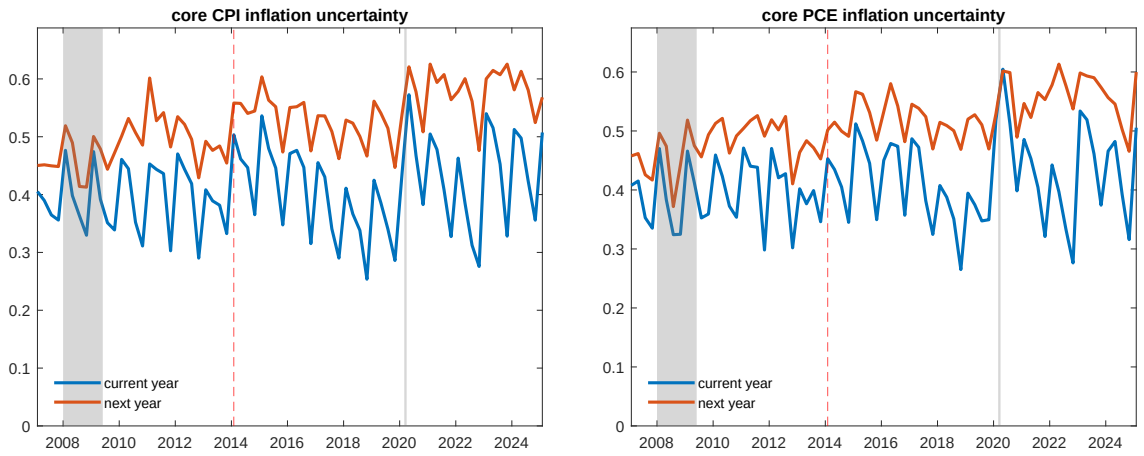

Figure 3: Aggregate uncertainty of core CPI inflation (left panel) and core PCE inflation (right panel) concerning the current and the next year. Uncertainty is measured by the standard deviation and averaged over forecasts for each survey round. The red vertical line indicates the date with change in bin width for GDP deflator growth. Shaded areas indicate recessions as dated by the NBER.

Our analysis focuses on three measures of inflation: GDP deflator growth, core CPI inflation, and core PCE inflation. The aim of our analysis is to demonstrate that the change in the bin width of GDP deflator inflation histograms in 2014 is the most likely explanation for the observed shift in measured uncertainty for this variable.

Breaks in inflation uncertainty from 2007 to 2025

As mentioned above, since 2007q1 the US SPF has also elicited histogram forecasts for core PCE inflation and core CPI inflation. No change in bin widths occurred for these variables in the sample period under study. As shown in Figure 3, the uncertainty concerning these inflation measures appears to be relatively stable around 2014q1.[8] Moreover, the evolution of realized quarterly inflation rates as displayed in Figure 4 does not indicate any major changes in the volatility of GDP deflator growth around the year 2014, and the same holds for core CPI inflation and core PCE inflation.

We investigate potential structural breaks in aggregate inflation uncertainty using the approach of Bai and Perron (1998, 2003). As a first step, we obtain seasonally-adjusted series by using the residuals of a regression with quarterly dummy variables. Denoting the uncertainty forecast for inflation measure π made in year t and quarter i for the year $t+j$ by $\sigma^{\pi}_{j|t,i}$, given a specific π and a specific j, we run the regression

$$\sigma^{\pi}_{j|t,i} = \sum_{q=1}^{4} \beta^{\pi}_{q,j} D_q + \varepsilon^{\pi}_{j;t,i} \tag{2}$$

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

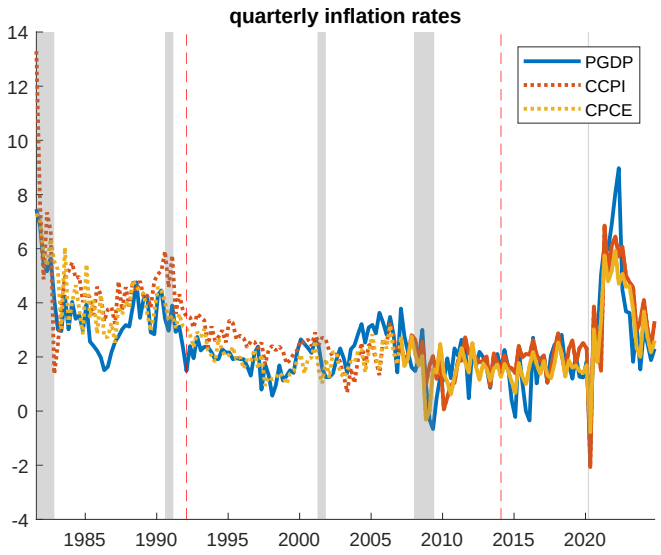

Figure 4: Annualized quarter-on-quarter inflation rates based on the seasonally adjusted GDP deflator (PGDP), core CPI (CCPI), and core PCE (CPCE) index from 1981q3 until 2025q1. Solid lines indicate the availability of corresponding histogram forecasts in the US SPF. The red vertical lines indicate dates with changes in bin widths for GDP deflator growth in the US SPF. Shaded areas indicate recessions as dated by the NBER.

with $\pi \in \{\text{GDP deflator growth, core CPI inflation, core PCE inflation}\}$, and $j = 0, 1$ denoting the forecast horizon in years (i.e. 0 for the current year, 1 for the next year). In each of these 6 equations, $i = 1, 2, 3, 4$ denotes the quarter the forecast is made, $t = 2007, 2008, \ldots, 2025$ denotes the year the forecast is made, and $\varepsilon^{\pi}_{j;t,i}$ denotes the residual. The dummy variables D_q are given by

$$D_q = \begin{cases} 0 & \text{if } q \neq i \\ 1 & \text{if } q = i. \end{cases} \tag{3}$$

The sample ranges from 2007q1 to 2025q1, i.e. it ends with $t = 2025$ and $i = 1$. The resulting residual series are displayed in Figure 11 in Appendix B.

We apply the sup $F_T(l+1|l)$ test of Bai and Perron (1998, 2003) to each of the six residual series obtained. That is, we sequentially test for $l+1$ versus l breaks starting with $l = 0$ and stopping when the test does not reject the null hypothesis of no additional break. We set the maximum number of structural breaks allowed to 5, use a significance level of 5%, and a trimming of 15 percent, corresponding to 11 observations. This trimming sample size enables the detection of breaks in samples from 2009q4 until 2022q2, while limiting size distortions.[9]

The results in Table 2 show that in 2014q1, breaks are detected only for GDP deflator growth uncertainty. They occurred for both horizons, and the break date is pinned down with

[9]Our codes are based on the MATLAB codes by Yohei Yamamoto available at https://blogs.bu.edu/perron/files/2024/12/m_break-matlab-YM.zip. We allow for all types of heterogeneity considered by Bai and Perron (1998, 2003), and for autocorrelation of the residuals. We use residual prewhitening and the repartition method of Bai (1997).

relatively high precision, resulting in short confidence intervals for both horizons. Breaks in inflation uncertainty for the next-year horizon are observed for core CPI and PCE inflation in 2009q4 and 2020q2 with lower precision. Assuming that the breaks in GDP deflator growth uncertainty in 2014q1 are caused by changes in histogram bin widths only, these results indicate a relatively stable inflation uncertainty from 2010 to 2019.

To corroborate these findings, in the spirit of McConnell and Perez-Quiros (2000), we analyze the residuals of first-order autoregressive models for the 3 inflation measures mentioned above based on realizations from 2006q4 to 2025q1. The absolute values of the 3 residual series obtained are shown in Figure 12 in Appendix B. They are tested with the $\sup F_T(l+1|l)$ test of Bai and Perron (1998, 2003) implemented as described above. The results displayed in Table 3 show that, for both core series, a structural break is found in 2020q2. While no break is detected in the GDP deflator growth volatility, the test finds a second break in core PCE inflation volatility in 2010q3. None of the break dates is estimated with high precision.

Based on these test results, and given the absence of events that could have strongly decreased GDP deflator growth uncertainty while leaving core inflation uncertainty constant around 2014q1, we conclude there is no evidence for a break in true inflation uncertainty around 2014q1. The test results suggest that most likely a break in uncertainty occurred around 2020, and that another break might have taken place about 10 years earlier.

Table 2: Structural breaks in aggregate inflation uncertainty

GDP deflator growth		
current year	2014q1 (7)	
next year	2014q1 (7)	
core CPI inflation		
current year	*no break detected*	
next year	2009q4 (16)	2020q2 (13)
core PCE inflation		
current year	*no break detected*	
next year	2009q4 (14)	2020q2 (20)

Note: Structural breaks are determined applying the sequential test of Bai and Perron (1998) to seasonally-adjusted current- and next-year uncertainty forecasts from equation (2). The numbers in parentheses denote the length of the respective 95% confidence interval around the estimated break date in quarters. See text for details.

So far, the results are based on simple seasonal adjustments of forecasts and on realizations. Moreover, they do not rely on tests using a known break date. Therefore, in what follows, we directly test the hypothesis of no change in uncertainty in 2014q1 based on the

Table 3: Structural breaks in inflation volatility

GDP deflator growth	*no break detected*	
core CPI inflation		2020q2 (20)
core PCE inflation	2010q3 (27)	2020q2 (24)

Note: Structural breaks are determined by applying the sequential test of Bai and Perron (1998, 2003) to the absolute values of the residuals of the regression $\pi_t = \rho\,\pi_{t-1} + \varepsilon_t$ estimated using samples from 2006q4 to 2025q1, yielding residuals starting in 2007q1. The numbers in parentheses denote the length of the respective 95% confidence interval around the estimated break date in quarters. See text for details.

(unadjusted) uncertainty forecasts for core CPI inflation and core PCE inflation. Moreover, we attempt to determine the sample with the least evidence for a break. This sample is likely to be informative about the effects of the bin width change on uncertainty, because the true uncertainty can be assumed to be virtually constant.

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

$$\sigma^{\pi}_{i,j|t} = c^{\pi}_{i,j} + \beta^{\pi}_{i,j}\, D_t + \varepsilon^{\pi}_{i,j;t} \tag{4}$$

with $\pi \in \{\text{core CPI inflation, core PCE inflation}\}$, $i = 1, 2, 3, 4$ denoting the quarter when 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), $t = 2007, 2008, \ldots, 2024$ denoting the year when the forecast is made, and $\varepsilon^{\pi}_{i,j;t}$ denoting the error term.[10] 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

$$D_t = \begin{cases} 0 & \text{if } t < 2014 \\ 1 & \text{if } t \geq 2014. \end{cases} \tag{5}$$

We test the null hypothesis $H_0 : \beta_{\pi,i,j} = 0$ for all π, i, j, i.e. for all 16 equations jointly. By varying the start and end years of the estimation window, we identify the sample that yields the largest p-value for H_0, implying the weakest statistical evidence against constant uncertainty across the two subsamples (i.e., the sample ending in 2013q4 and the sample starting in 2014q1). We require the samples to have at least two years of observations before and after the potential structural break in 2014, corresponding to at least two observations to estimate each parameter in equation (4).

[10] We do not use the forecasts from 2025q1, because we want to obtain a balanced sample, i.e. we want to have the same number of observations for each i.

From Table 4, we can see that the null hypothesis of constant uncertainty is always rejected in samples starting before 2009, or ending in 2024 at the 5% significance level. In general, if one of the subsamples considered for the estimation is very short, the null hypothesis can often be rejected. This is most likely due to cyclical variations or single shocks rather than to structural change. As mentioned above, the parameters in equation (4) are estimated on a relatively small number of observations in short subsamples, such that those observations being close to a cyclical peak or a trough, or the occurrence of a single large shock can perceptibly affect the results.[11] The largest p-value is obtained for the sample from 2010 to 2019. This sample also has the lowest average of the estimates of $\beta^{\pi}_{i,j}$ from (4), being equal to 0.011. This number appears relatively small when considering the values of the dependent variables displayed in Figure 3.

In summary, it seems reasonable to assume that inflation forecast uncertainty from 2010 to 2013 is broadly equal to inflation forecast uncertainty from 2014 to 2019. We proceed under the assumption of equality. This will allow us to gauge the effect of the bin width change in 2014 on *measured* inflation forecast uncertainty.

Table 4: Tests for change in forecast uncertainty in 2014 for core CPI and core PCE inflation

end:	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
start	p-values of F-test of $H_0 : \beta_{\pi,i,j} = 0$ for all π, i, j jointly									
2007	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2008	0.00	0.00	0.00	0.01	0.03	0.00	0.00	0.00	0.00	0.00
2009	0.00	0.00	0.00	**0.14**	**0.32**	**0.07**	0.01	0.01	0.00	0.00
2010	0.00	0.00	0.01	**0.50**	***0.74***	**0.40**	**0.15**	**0.11**	0.02	0.01
2011	0.01	0.00	0.01	**0.50**	**0.66**	**0.49**	**0.23**	**0.18**	**0.05**	0.02
2012	0.00	0.00	0.00	**0.18**	**0.32**	**0.28**	**0.10**	**0.08**	0.03	0.01
start	average of estimates of $\beta^{\pi}_{i,j}$ from (4) over all π, i, j									
2007	0.05	0.05	0.04	0.03	0.03	0.03	0.04	0.04	0.04	0.04
2008	0.05	0.04	0.04	0.03	0.02	0.03	0.03	0.03	0.04	0.04
2009	0.04	0.04	0.03	**0.02**	**0.02**	**0.03**	0.03	0.03	0.03	0.03
2010	0.04	0.04	0.03	**0.02**	***0.01***	**0.02**	**0.03**	**0.03**	0.03	0.03
2011	0.04	0.04	0.03	**0.02**	**0.02**	**0.02**	**0.03**	**0.03**	**0.03**	0.03
2012	0.05	0.05	0.04	**0.03**	**0.03**	**0.03**	**0.04**	**0.04**	0.04	0.04

Note: 'start' denotes first year, 'end' last year of OLS estimation sample. Bold numbers indicate p-values larger than 0.05. Numbers in italics show largest p-value, and smallest average of estimates of $\beta^{\pi}_{i,j}$, respectively.

Gauging the effect of halving bin widths on measured uncertainty

To quantify the effect of a bin width change on measured uncertainty, we assume that there is an elasticity γ which describes how a change in the coarseness of the histogram x affects

[11] Figure 11 in Appendix B suggests that a typical cycle in inflation forecast uncertainty, gauged as the time between two peaks or two troughs, appears to last longer than 3 years.

the change in measured uncertainty σ_{meas}, i.e., we assume a relation given by

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

where our measure of coarseness x simply refers to pp per histogram bin. We choose this specification because it relies on a single parameter, which can be determined if only one bin width change is observed. Moreover, it is easy to interpret, and it 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. The results above suggest that this assumption does not hold.

For the following expositions, it is sometimes helpful to consider the equation for the level of measured uncertainty implied by equation (6). This is given by

$$\sigma_{meas} = c \, x^{\gamma}. \tag{7}$$

Glas and Hartmann (2022) consider a thought experiment corresponding to $\gamma = 1$, i.e. to a situation where respondents always assign positive probabilities to a fixed number of bins only. This implies that they ignore the bin widths when making their histogram forecasts.[12] 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 mean 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 consider the regression corresponding to (7), namely

$$\ln \left(\sigma^{\pi}_{j|t,i} \right) = c_{i,j} + \gamma_{i,j} \, \ln \left(x^{\pi}_{t,i} \right) + \varepsilon_{t,i,j} \tag{8}$$

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

$$x^{\pi}_{t,i} = \begin{cases} 1 & \text{if } t < 2014 \\ 0.5 & \text{if } t \geq 2014, \end{cases} \tag{9}$$

which corresponds to the histogram bin widths in pp.[13]

We estimate equation (8) for the years 2010 to 2019 because inflation forecast uncertainty in the sample from 2010 to 2013 is approximately equal to inflation forecast uncertainty from 2014 to 2019 according to our previous results. Therefore, the coefficients $\gamma_{i,j}$ can be interpreted as capturing the effect of the change in bin width on measured forecast

[12] Glas and Hartmann (2022) actually use the variance to measure uncertainty. If we square both sides of (7), 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 $\gamma = 1$.

[13] 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 can be considered inconsequential in the definition of $x^{\pi}_{t,i}$.

Table 5: Horizon-dependent effects of bin width change on measured uncertainty

$c_{1,1}$	$c_{2,1}$	$c_{3,1}$	$c_{4,1}$	$c_{1,0}$	$c_{2,0}$	$c_{3,0}$	$c_{4,0}$
-0.240	-0.283	-0.279	-0.336	-0.399	-0.443	-0.465	-0.684
(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)
$\gamma_{1,1}$	$\gamma_{2,1}$	$\gamma_{3,1}$	$\gamma_{4,1}$	$\gamma_{1,0}$	$\gamma_{2,0}$	$\gamma_{3,0}$	$\gamma_{4,0}$
0.512	0.488	0.527	0.517	0.563	0.567	0.674	0.667
(0.080)	(0.080)	(0.080)	(0.080)	(0.080)	(0.080)	(0.080)	(0.080)

Note: Estimates of $c_{i,j}$ and $\gamma_{i,j}$ from equation (8) for the sample 2010 to 2019. Forecast horizon decreases from $i = 1, j = 1$ (forecasts from 1st quarter for next year) to $i = 4, j = 0$ (nowcast from 4th quarter for current year). Numbers in parentheses are OLS standard errors. They are identical for all $c_{i,j}$ resp. $\gamma_{i,j}$ estimates due to the special form of the regressor matrix.

uncertainty.

Table 5 shows that, as expected, the coefficients $c_{i,j}$ decrease with the forecast horizon. The estimates of the coefficients $\gamma_{i,j}$ are mostly around 0.5 to 0.6 for all 8 forecast horizons. While the current-year coefficients $\gamma_{i,0}$ appear to be marginally larger than their next-year counterparts $\gamma_{i,1}$, running an F-test of the null hypothesis that all $\gamma_{i,j}$ are equal yields a p-value of 0.60. Therefore, we estimate equation (8) with the restriction $\gamma_{i,j} = \gamma$ imposed for all i and j. We do so for all samples where the null hypothesis of no structural break in inflation uncertainty in 2014q1 cannot be rejected according to the results in Table 4.

The estimation results in Table 6 show that the estimates of γ vary between about 0.5 and 0.6 depending on the sample. The smallest values are observed if the estimation sample starts in 2012, such that only two years of data are available before the change in the bin width.[14] Our preferred sample from 2010 to 2019 yields $\gamma \approx 0.56$. Looking at the standard errors of the estimates, one can reject both extreme assumptions mentioned concerning the effect of a change in the histogram bin width on measured uncertainty. In all samples, the usual assumption of $\gamma = 0$ has a p-value of virtually zero, implying that measured uncertainty does change in response to the change in bin width. The assumption of $\gamma = 1$ 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 results show that an intermediate assumption is warranted. This aggregate result could, in principle, be caused by the presence of two groups of forecasters with $\gamma \approx 0$ and $\gamma \approx 1$, respectively. However, when we estimate individual elasticities, they rather appear to have a hump-shaped distribution around their average of about 0.6. More details on this estimation can be found in Appendix C.

With $\gamma = 0.56$, halving the bin width yields a measured standard deviation that is smaller by a factor of 0.68, i.e. it causes the measured standard deviation to decrease by about one-third. This result simply follows from $\gamma \ln(1/2) \approx -0.39$ and $\exp(-0.39) \approx 0.68$.[15]

[14] The null hypothesis $\gamma_{i,j} = \gamma$ in the corresponding estimations with horizon-dependent values $\gamma_{i,j}$ cannot be rejected for any of the samples considered. Yet, the power of these tests is likely to be small, especially in the shorter samples.

[15] Based on a household survey, 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

Table 6: Effects of bin width change on measured uncertainty under horizon independence

	sample end					
	2018	2019	2020	2021	2022	2023
sample start						
2009	0.572	0.581	0.571			
	(0.027)	(0.026)	(0.025)			
2010	0.556	**0.564**	0.555	0.541	0.560	
	(0.029)	**(0.028)**	(0.027)	(0.026)	(0.030)	
2011	0.554	0.563	0.553	0.539	0.558	0.539
	(0.032)	(0.031)	(0.030)	(0.029)	(0.034)	(0.034)
2012	0.487	0.495	0.486	0.472	0.491	
	(0.031)	(0.031)	(0.031)	(0.031)	(0.037)	

Note: Estimates of γ from equation (8) with restriction $\gamma_{i,j} = \gamma$ imposed. Estimates are shown for samples without breaks according to results in Table 4. Numbers in parentheses are OLS standard errors. The bold numbers indicate the sample with the highest p-value in Table 4.

3 Adjusting SPF forecast uncertainty for breaks

Based on equation (7), we can determine a break-adjusted GDP deflator growth uncertainty $\tilde{\sigma}^{\pi}_{j|t,i}$ as

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

with $\gamma = 0.56$. Since $x^{\pi}_{t,i}$, the bin width in pp, equals 1 from 1992q1 to 2013q4, this break-adjusted uncertainty coincides with the original uncertainty over that 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, 1 from 1992q1 to 2013q4, and 0.5 from 2014q1 to 2025q1.

This type of break adjustment, however, cannot directly be applied to real GDP growth uncertainty since 2020q2. This is because from that date onward many different bin widths coexist, 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 changed neither in 2020q2 nor in 2024q2.

Denoting the aggregate real GDP growth uncertainty obtained in this way for each quar-

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. Yet, their results clearly do not support the assumption $\gamma = 0$, either.

ter since 2020q2 by $\acute{\sigma}^{y}_{j|t,i}$, we can exploit the relation

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

from 2020q2 onwards. 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 $\acute{\sigma}^{y}_{j|t,i}$ is calculated corresponds to the assumption $\gamma = 1$. In this case, the expected measured uncertainty for any hypothetical bin definition (having the same number of bins as in the original setting) could be found simply by applying this bin definition to the probabilities provided. Correspondingly, (11) follows from (6) with $\gamma = 1$. Since the resulting coarseness depends on the probabilities of the histograms, it also depends on the forecast horizon j — in contrast to the episodes with constant bin widths.

Figure 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 bin width change in 2020q2. 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. In 2024q2, the coarseness hardly changes, although the width of 8 out of 11 bins is halved. This is because about 90% of the probability mass from 2024q2 to 2025q1 is allocated to the central three bins whose width remains unchanged.

Assuming that $\gamma = 0.56$ is a reasonable elasticity for the uncertainty forecasts of real GDP growth as well, we proceed by using equation (10) to obtain its break-adjusted uncertainty series. That is, we calculate the break-adjusted series $\tilde{\sigma}^{y}_{j|t,i}$ as

$$\tilde{\sigma}^{y}_{j|t,i} = \sigma^{y}_{j|t,i} \left(x^{y}_{t,i,j}\right)^{-\gamma} \tag{12}$$

with $\gamma = 0.56$. Again, $x^{\pi}_{t,i,j}$ equals 2 from the start of the sample until 1991q4. Similar to GDP deflator growth, we have $x^{y}_{t,i,j} = 1$ from 1992q1 to 2020q1, such that the break-adjusted uncertainty coincides with the original uncertainty over this period. From 2020q2 onward, $x^{y}_{t,i,j}$ changes each period and depends on the forecast target j.

Using the coarseness series in equations (10) and (12) finally gives the break-adjusted series, which are displayed in Figure 6. While the structural break in measured uncertainty of GDP deflator growth in 2014q1 disappears by construction, the lack of visible structural breaks in 1992q1 for GDP deflator and real GDP growth is remarkable. For GDP deflator growth, uncertainty appears at best marginally larger before 1992q1 than in the years afterward. Since 2020, GDP deflator growth uncertainty is above average. This conclusion stands in stark contrast to the impression conveyed by the unadjusted uncertainty series in Figure 1. For real GDP growth, uncertainty seems to remain relatively constant from the

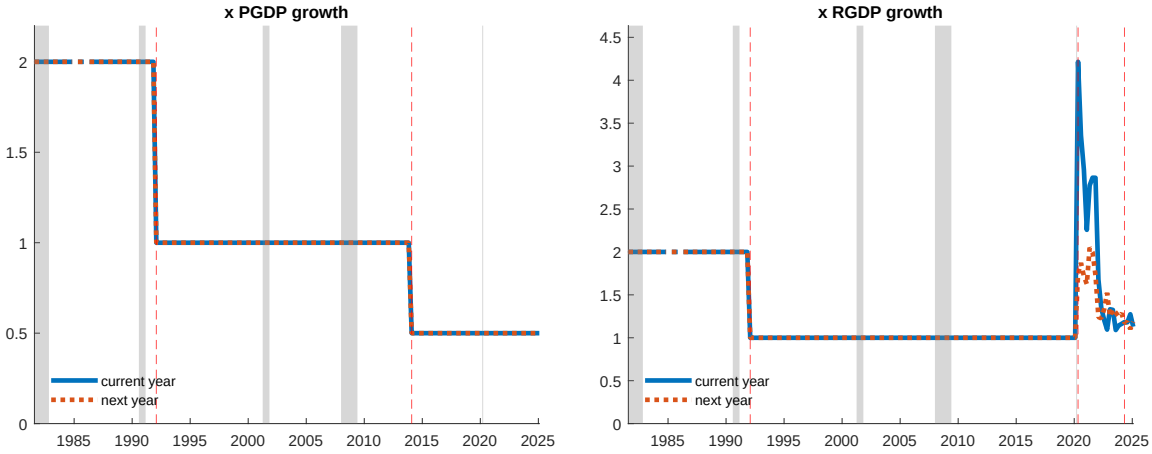

Figure 5: Coarseness of histograms as measured by $x^y_{t,i,j} = x^\pi_{t,i}$ for GDP deflator growth (left panel) and $x^y_{t,i,j}$ for real GDP growth (right panel). Red vertical lines indicate the dates with changes in bin widths. Shaded areas indicate recessions as dated by the NBER.

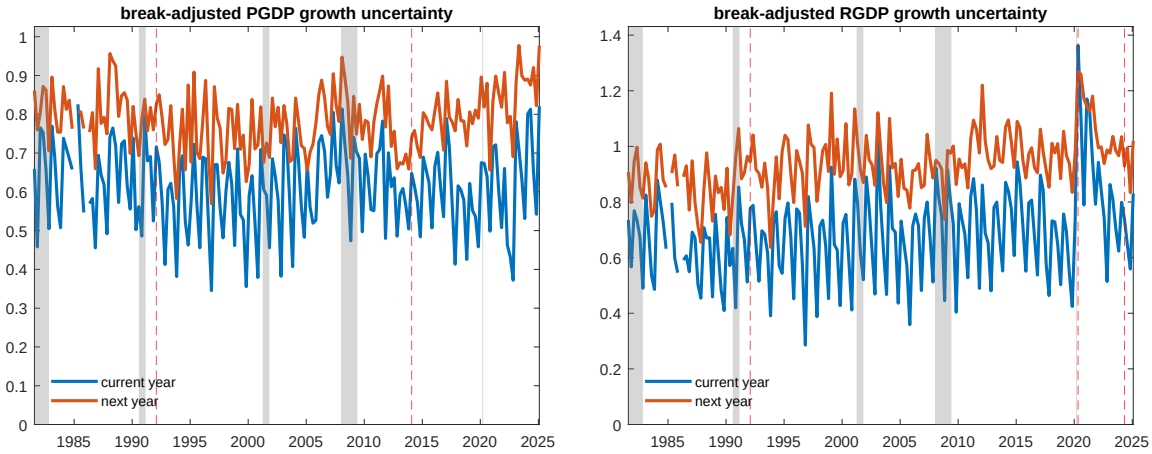

Figure 6: Break-adjusted uncertainty of GDP deflator growth (left panel) and of real GDP growth (right panel). Red vertical lines indicate the dates with changes in bin widths. Shaded areas indicate recessions as dated by the NBER.

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 the increase is considerably less drastic than conveyed by the unadjusted uncertainty series in Figure 1. Like the unadjusted series, growth uncertainty at the end of the sample broadly equals its pre-pandemic level.

A notable feature that becomes more evident in the break-adjusted uncertainty series, although it is also present in the unadjusted series, is the lack of a 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 real GDP growth, only the pandemic recession is clearly associated with higher uncertainty. For GDP deflator growth, 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 6. On some occasions, high uncertainty might be masked by the expected realizations being close to the open bins. For instance, the relatively low uncertainty concerning GDP deflator growth in 2022 mainly stems from the fact that high probabilities are assigned to the upper open bin. Therefore, the uncertainty estimated for the corresponding surveys strongly depends on the researchers' assumptions concerning the probability distribution in the open bins.

4 A modest proposal for the US SPF's bin widths

In light of the previous results, one might consider modifications to the way probabilistic forecasts are elicited. While ideas like using questions for quantiles, five-point distributions as in Altig et al. (2022), or choice-based histograms as in Goldfayn-Frank et al. (2025) are promising, they would imply a profound change in the way the US SPF is conducted. If the US SPF continues to rely on its traditional histogram forecasts, it would seem reasonable to reconsider the choice of the bin widths due to their impact on outcomes. If changes in the bin widths lead to breaks in measured uncertainty, one natural question to ask is, how a constant bin width should have been chosen in our sample. This bin width would also be a choice worth considering for future survey rounds, unless the average uncertainty in the coming decades differs noticeably from the average uncertainty in our sample.

The literature on forecast surveys often states that an uncertainty forecast should be unbiased in the sense that it corresponds to the expectation of a measure of ex post uncertainty. Such a relation is relatively straightforward to investigate for an individual forecast. For instance, an unbiased variance forecast of forecaster n implies

$$\left(\sigma^z_{n;j|t,i}\right)^2 = \mathbb{E}\left[\left(z_{t+j,i} - \mu^z_{n;j|t,i}\right)^2\right], \tag{13}$$

where the right-hand side is the expected squared forecast error, which is the measure of ex post uncertainty corresponding to the variance forecast. $\mu^z_{n;j|t,i}$ denotes the mean fore-

cast from the same forecast distribution that $\sigma^z_{n;j|t,i}$ is derived from, and $z_{t+j,i}$ denotes the realization of the target variable. In the previous sections, we found that a change in the bin width affects the left-hand side of equation (13). In this section, we assume that such a change does not affect the mean forecast $\mu^z_{n;j|t,i}$, i.e. the right-hand side of equation (13). This assumption does not appear particularly strong, because forecasters tend to align the central tendency resulting from their histogram forecasters with their point predictions (see Engelberg et al. 2009).

Since we are focusing on squared aggregate uncertainty, we employ the aggregate equality implied by (13), which is given by

$$\frac{1}{N_{i,t}}\sum_{n=1}^{N_{i,t}}\left(\sigma^z_{n;j|t,i}\right)^2 = \mathbb{E}\left[\frac{1}{N_{i,t}}\sum_{n=1}^{N_{i,t}}\left(z_{t+j,i}-\mu^z_{n;j|t,i}\right)^2\right]. \tag{14}$$

Note that this equality does not require every single forecaster to issue unbiased variance forecasts. Yet, the average bias equals zero if (14) holds. Defining the combined mean forecast, i.e. the average mean forecast across all forecasters by

$$\overline{\mu}^z_{j|t,i} = \frac{1}{N_{i,t}}\sum_{n=1}^{N_{i,t}}\mu^z_{n;j|t,i}$$

the right-hand side of (14) allows the useful decomposition

$$\begin{aligned}\mathbb{E}\left[\frac{1}{N_{i,t}}\sum_{n=1}^{N_{i,t}}\left(z_{t+j,i}-\mu^z_{n;j|t,i}\right)^2\right] &= \mathbb{E}\left[\frac{1}{N_{i,t}}\sum_{n=1}^{N_{i,t}}\left(\left(z_{t+j,i}-\overline{\mu}^z_{j|t,i}\right)-\left(\mu^z_{n;j|t,i}-\overline{\mu}^z_{j|t,i}\right)\right)^2\right]\\ &= \mathbb{E}\left[\left(z_{t+j,i}-\overline{\mu}^z_{j|t,i}\right)^2\right]+\mathbb{E}\left[\frac{1}{N_{i,t}}\sum_{n=1}^{N_{i,t}}\left(\mu^z_{n;j|t,i}-\overline{\mu}^z_{j|t,i}\right)^2\right].\end{aligned} \tag{15}$$

Combining (14) and (15) yields

$$\frac{1}{N_{i,t}}\sum_{n=1}^{N_{i,t}}\left(\sigma^z_{n;j|t,i}\right)^2 = \mathbb{E}\left[\left(z_{t+j,i}-\overline{\mu}^z_{j|t,i}\right)^2\right]+\mathbb{E}\left[\frac{1}{N_{i,t}}\sum_{n=1}^{N_{i,t}}\left(\mu^z_{n;j|t,i}-\overline{\mu}^z_{j|t,i}\right)^2\right]. \tag{16}$$

Equation (16) shows that, under variance unbiasedness, the average variance forecast equals the expected squared error of the combined mean forecast (i.e., the expectation of an ex post uncertainty measure) plus a term capturing disagreement (an ex ante uncertainty measure). The disagreement term equals the average squared difference between the individual mean forecasts and the combined mean forecast. The equality in (16) holds for each single forecast, i.e. for each (j,t,i).[16]

In what follows, it will be useful to consider the forecast horizon in quarters denoted by h,

[16]The result stated in (16) is directly related to the upward bias of the variance forecasts of linear pools, as discussed in Proposition 1 in Knüppel and Krüger (2022).

where $h = 0$ corresponds to $(i = 4, j = 0)$ i.e., to the current-year forecast made in the fourth quarter, $h = 1$ to $(i = 3, j = 0)$, ..., and $h = 7$ to $(i = 1, j = 1)$ i.e., to the next-year forecast made in the first quarter. That is, a larger h corresponds to a longer forecast horizon in quarters. Using the definition of aggregate uncertainty from (1) and replacing expectations with sample averages in (16) yields

$$\frac{1}{T}\sum_{t=1}^{T}\underbrace{\left(\sigma^z_{j|t,i}\right)^2}_{\sigma^2_{h,t}} = \frac{1}{T}\sum_{t=1}^{T}\underbrace{\left(z_{t+j,i} - \overline{\mu}^z_{j|t,i}\right)^2}_{S_{h,t}} + \frac{1}{T}\sum_{t=1}^{T}\underbrace{\left(\frac{1}{N_{i,t}}\sum_{n=1}^{N_{i,t}}\left(\mu^z_{n;j|t,i} - \overline{\mu}^z_{j|t,i}\right)^2\right)}_{D_{h,t}} + o_p(1), \tag{17}$$

assuming covariance stationarity, ergodicity, and finite first moments of $\sigma^2_{h,t}$, $S_{h,t}$ and $D_{h,t}$. $\sigma^2_{h,t}$ denotes the average aggregate squared uncertainty for forecast horizon h in year t. In finite samples with unbiased variance forecasts, the sample average of $\sigma^2_{h,t}$ becomes approximately equal to the sum of the sample averages of the ex post uncertainty $S_{h,t}$ and disagreement $D_{h,t}$ for forecast horizon h as T grows. For the quantities mentioned and in what follows, we do not use the index z to avoid clutter.

In our samples, $\sigma^2_{h,t}$, $S_{h,t}$, and $D_{h,t}$ have missing values corresponding to the forecasts from 1984q1 and 1985q1. Since we are going to investigate averages for each h, two missing values for $h = 3$ and $h = 7$, respectively, might distort the results, not least since these missing values occurred in potentially more volatile times. Therefore, we employ fitted values of regressions to fill these gaps. We use balanced samples defined with respect to the target year of the forecasts. Our samples begin in 1983q1, implying that they start with the forecasts made in 1983q1 for the current year ($h = 3$) and forecasts made in 1982q1 for the next year ($h = 7$). Looking at the forecasts *for* instead of *from* a certain year is motivated by the fact that the largest quantities in (17) are observed for $S_{h,t}$, and these occur in years with extreme realizations. Our samples for the interpolation regressions end in 2019 because, otherwise, the year 2020, when the COVID-19 pandemic started, would strongly affect the results for real GDP growth. By ending the samples in 2019, we also exclude the extreme GDP deflator growth of the year 2022.

We consider the unadjusted aggregate squared uncertainty $\sigma^2_{h,t}$, and the break-adjusted aggregate squared uncertainty $\tilde{\sigma}^2_{h,t}$. To obtain fitted values, we regress $\sigma^2_{3,t}$ on $\sigma^2_{2,t}$ and on $\sigma^2_{4,t-1}$, while $\sigma^2_{7,t}$ is regressed on $\sigma^2_{6,t}$ only.[17] We do the same for $\tilde{\sigma}^2_{h,t}$, $S_{h,t}$ and $D_{h,t}$. Figure 7 shows the time series with the fitted values. The series $\sigma^2_{h,t}$ ($\tilde{\sigma}^2_{h,t}$) in Figure 7 is closely related to the uncertainty series in Figure 6 (Figure 1). In contrast to the latter, the former contains variances instead of standard deviations, and missing values are filled with fitted values. Moreover, the values of the next-year forecasts displayed are shifted by 4 quarters to the right because of looking at forecasts *for* a certain year. The ex post measure $S_{h,t}$ has spikes with very large values in years when the outturns are extreme. Disagreement $D_{h,t}$ tends to be more stable over time, and its maxima are far smaller than those of $S_{h,t}$.

[17] We do not include a constant as a regressor because we assume that the uncertainties for adjacent horizons are proportional to each other.

Based on these series, we can investigate whether (17) holds at least approximately for one of the measures of squared forecast uncertainty. To this end, we define the sample averages

$$\begin{aligned}
\overline{\sigma}_h^2 &= \frac{1}{T'-t'+1}\sum_{t=t'}^{T'}\sigma_{h,t}^2 \\
\overline{\tilde{\sigma}}_h^2 &= \frac{1}{T'-t'+1}\sum_{t=t'}^{T'}\tilde{\sigma}_{h,t}^2 \\
\overline{S}_h &= \frac{1}{T'-t'+1}\sum_{t=t'}^{T'}S_{h,t} \\
\overline{D}_h &= \frac{1}{T'-t'+1}\sum_{t=t'}^{T'}D_{h,t}.
\end{aligned} \tag{18}$$

with $t'=1, T'=T$ if $h \leq 3$ (current-year forecasts) and $t'=0, T'=T-1$ if $h \geq 4$ (next-year forecasts). The sample to calculate the averages starts with the forecasts for 1983 and ends with the forecasts for 2024.

Table 7 shows the values of the sample averages of ex post uncertainty $\overline{S}_h$ and disagreement $\overline{D}_h$ for the full samples and the samples without the years with extreme observations (2020 for real GDP growth, 2022 for GDP deflator growth). In all samples, except for short forecast horizons, $\overline{S}_h$ exceeds $\overline{D}_h$, considerably so for large horizons. Excluding extreme observations strongly reduces $\overline{S}_h$ except for very short horizons for RGDP growth, whereas $\overline{D}_h$ remains broadly constant.[18]

Against the background of the extreme errors of the real GDP growth forecasts for 2020 and the GDP deflator growth forecasts for 2022, in what follows, we exclude the forecasts for these years when calculating the averages. An additional motivation for the latter exclusion is that for the current-year forecasts, large probabilities were assigned to the upper outer bin, such that forecast variance assessments for these forecasts strongly depend on assumptions. However, we provide results for the full sample in Appendix D.

In line with the previous sections, we now apply square roots to the quantities defined in (18) when describing uncertainty. Note that in the previous sections, we looked at square roots of cross-sectional variances, whereas here we have square roots of time-series variances. $\overline{\sigma}_h := \sqrt{\overline{\sigma}_h^2}$ and $\overline{\tilde{\sigma}}_h := \sqrt{\overline{\tilde{\sigma}}_h^2}$ denote the unadjusted and break-adjusted *average forecast uncertainty*,[19] respectively, while the quantity $\sqrt{\overline{S}_h + \overline{D}_h}$ denotes *calibration uncertainty*, because the average of unbiased aggregate uncertainty should be approximately equal to this term according to (17).

As Figure 8 shows, neither $\overline{\sigma}_h$ nor $\overline{\tilde{\sigma}}_h$ approximate the calibration uncertainties well. For both variables, the average forecast uncertainties are too small for forecast horizons $h \geq 3$.

[18] Regarding the next-year forecasts ($h = 7, 6, 5, 4$), one could suppose that this behavior of $\overline{D}_h$ is partly due to focusing on forecasts for (not from) 2022 and 2020, respectively. Yet, this result also emerges when excluding forecasts from 2022 and 2020.

[19] Note that $\overline{\sigma}_h$ and $\overline{\tilde{\sigma}}_h$ are not the sample averages of standard deviations.

Table 7: Sample averages of ex post uncertainty $\overline{S}_h$ and disagreement $\overline{D}_h$

	PGDP growth				RGDP growth			
	$\overline{S}_h$		$\overline{D}_h$		$\overline{S}_h$		$\overline{D}_h$	
sample	full	w/o 2022	full	w/o 2022	full	w/o 2020	full	w/o 2020
$h = 7$	1.62	1.06	0.43	0.44	3.00	2.37	0.56	0.56
$h = 6$	1.48	0.98	0.41	0.41	2.62	1.96	0.61	0.61
$h = 5$	1.22	0.82	0.31	0.30	2.41	1.79	0.49	0.49
$h = 4$	0.84	0.51	0.30	0.30	1.98	1.36	0.43	0.44
$h = 3$	0.57	0.34	0.31	0.30	1.44	0.77	0.43	0.44
$h = 2$	0.41	0.23	0.24	0.24	0.45	0.31	0.34	0.25
$h = 1$	0.27	0.09	0.20	0.20	0.17	0.12	0.25	0.20
$h = 0$	0.25	0.08	0.20	0.19	0.10	0.10	0.24	0.21

Note: Averages of samples starting with forecasts for 1983 and ending with the forecasts for 2024. For PGDP growth, the sample w/o 2022 excludes the forecasts for 2022. For RGDP growth, the sample w/o 2020 excludes the forecasts for 2020. h denotes the forecast horizon in quarters.

The average unadjusted uncertainties run virtually parallel to the average break-adjusted uncertainties, but are slightly larger. The corresponding results for the full sample can be found in Figure 14 in Appendix D.

For the break-adjusted uncertainties, we chose a coarseness of $x = 1$, corresponding to a bin width of 1 pp. While we could employ a larger value of x, this would only amount to multiplying the break-adjusted forecast uncertainty $\overline{\tilde{\sigma}}_h$ by the same constant (larger than 1) for each h. Having larger forecast uncertainties at larger horizons would then lead to a mismatch at shorter horizons.

To attain break-adjusted forecast uncertainties that match the calibration uncertainties for all h, one needs to consider horizon-specific bin widths. Denoting these bin widths by x_h^*, the corresponding time series of aggregate uncertainty and the average forecast uncertainty are given by

$$\begin{aligned} \tilde{\sigma}_{h,t}^* &= \tilde{\sigma}_{h,t} \, (x_h^*)^\gamma \\ \overline{\tilde{\sigma}}_h^* &= \sqrt{\frac{1}{T' - t' + 1} \sum_{t=t'}^{T'} \left(\tilde{\sigma}_{h,t}^* \right)^2} \end{aligned} \tag{19}$$

where x_h^* is chosen such that

$$\overline{\tilde{\sigma}}_h^* = \sqrt{\overline{S}_h + \overline{D}_h} \tag{20}$$

holds. Note that (20) does not imply that each forecaster makes unbiased variance forecasts as defined in (13). However, it implies that the individual biases add up to zero. Put differently, when randomly choosing a forecaster, the expected bias of this forecaster's variance forecast is zero if the bin width x_h^* is used.

While advocating the use of x_h^* for the US SPF might appear straightforward, two practical issues warrant attention. First, US SPF bin widths have always been multiples of 0.5 pp, and other choices could be difficult to handle. Second, it might be desirable to have

monotonicity of x_h^* with respect to h. Figure 8 suggests that, in general, bin widths should decrease with h because the ratio of calibration uncertainty to average uncertainty decreases with h. Yet, x_0^* would need to be larger than x_1^* for both average forecast uncertainties, $\overline{\widetilde{\sigma}}_0^*$ and $\overline{\widetilde{\sigma}}_1^*$, to match their respective calibration uncertainties. However, survey participants might be irritated by an increase in the bin width for the current-year forecasts from the third to the fourth quarter if, at the same time, the bin width for the next-year forecasts decreases, and, moreover, if the bin widths for both forecasts have not increased before.

The first issue can be resolved by rounding each x_h^* to the nearest multiple of 0.5. We denote these rounded bin widths by $r\left(x_h^*\right)$. The second issue can be resolved by running an isotonic regression using the Pool Adjacent Violators Algorithm (PAVA). This regression finds the values $\widehat{x}_h^*$ which minimize

$$\sum_{h=0}^{7} (\widehat{x}_h^* - x_h^*)^2 \tag{21}$$

subject to the constraints $\widehat{x}_h^* \leq \widehat{x}_{h+1}^*$ for $h = 0, 1, \ldots, 6$. The rounded fitted values will be denoted by $r\left(\widehat{x}_h^*\right)$. Note that we round after imposing monotonicity, not vice versa.

Table 8: Calibration-uncertainty-based bin widths

	PGDP growth				RGDP growth			
	x_h^*	$\widehat{x}_h^*$	$r\left(x_h^*\right)$	$r\left(\widehat{x}_h^*\right)$	x_h^*	$\widehat{x}_h^*$	$r\left(x_h^*\right)$	$r\left(\widehat{x}_h^*\right)$
$h=7$	2.05	2.05	2.00	2.00	2.65	2.65	2.50	2.50
$h=6$	1.96	1.96	2.00	2.00	2.41	2.41	2.50	2.50
$h=5$	1.73	1.73	1.50	1.50	2.35	2.35	2.50	2.50
$h=4$	1.46	1.46	1.50	1.50	2.16	2.16	2.00	2.00
$h=3$	1.27	1.27	1.50	1.50	1.66	1.66	1.50	1.50
$h=2$	1.02	1.02	1.00	1.00	0.96	0.99	1.00	1.00
$h=1$	0.82	0.97	1.00	1.00	0.76	0.99	1.00	1.00
$h=0$	1.12	0.97	1.00	1.00	1.25	0.99	1.50	1.00

Note: US SPF bin widths (in pp) making the average forecast uncertainty (approximately) equal to the calibration uncertainty for the sample 1983 to 2024, excluding 2022 for GDP deflator growth and 2020 for real GDP growth. x_h^* leads to equality for all h. $\widehat{x}_h^*$ enforces monotonicity of the bin widths with respect to h. $r(\bullet)$ rounds both types of bin widths to multiples of 0.5.

The resulting bin widths in Table 8 show values in the range of about 1.0 pp to 2.5 pp. The bin widths tend to be larger for real GDP growth, especially for $h \geq 3$. As mentioned above, the bin widths x_h^* decrease with h except for $h = 0$. Therefore, the isotonic regressions yield equal bin widths $\widehat{x}_h^*$ for $h = 0, 1$ for GDP deflator growth, and for $h = 0, 1, 2$ for real GDP growth, respectively. While the differences between x_h^* and $\widehat{x}_h^*$ vanish after rounding to multiples of 0.5, the results differ for the corresponding rounded values for real GDP growth at $h = 0$.

We will focus on the bin widths $r\left(\widehat{x}_h^*\right)$, because they account for both practical issues mentioned above. While these bin widths might appear large for large h, they are rounded downwards for $h = 7$. Moreover, if the full samples, i.e. the samples from 1983 to 2024

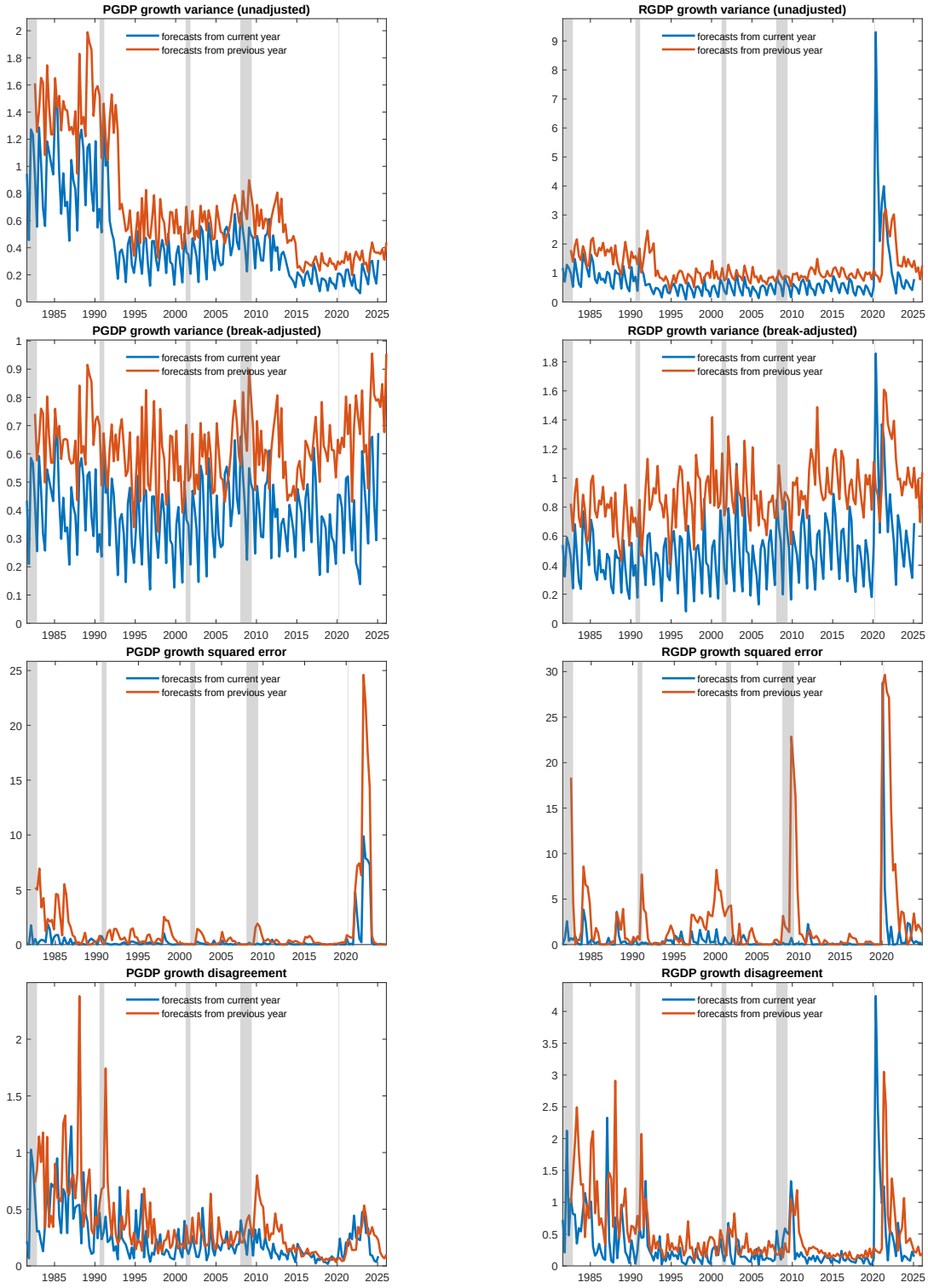

Figure 7: From the top to the bottom, $\sigma^2_{h,t}$, $\tilde{\sigma}^2_{h,t}$, $S_{h,t}$, $D_{h,t}$ as described for equation (17) are displayed for GDP deflator growth (left panel) and real GDP growth (right panel). Blue lines show forecasts with $h = 0, 1, 2, 3$, red lines show forecasts with $h = 4, 5, 6, 7$. So, given $h = 0, 1, 2, 3$, $\sigma^2_{h,t}$ $(\tilde{V}_{h,t}, S_{h,t}, D_{h,t})$ and $\sigma^2_{h+4,t-1}$ $(\tilde{\sigma}^2_{h+4,t-1}, S_{h+4,t-1}, D_{h+4,t-1})$ are plotted together, i.e. the values for forecasts *for* instead of *from* a certain year on the x-axis are shown.

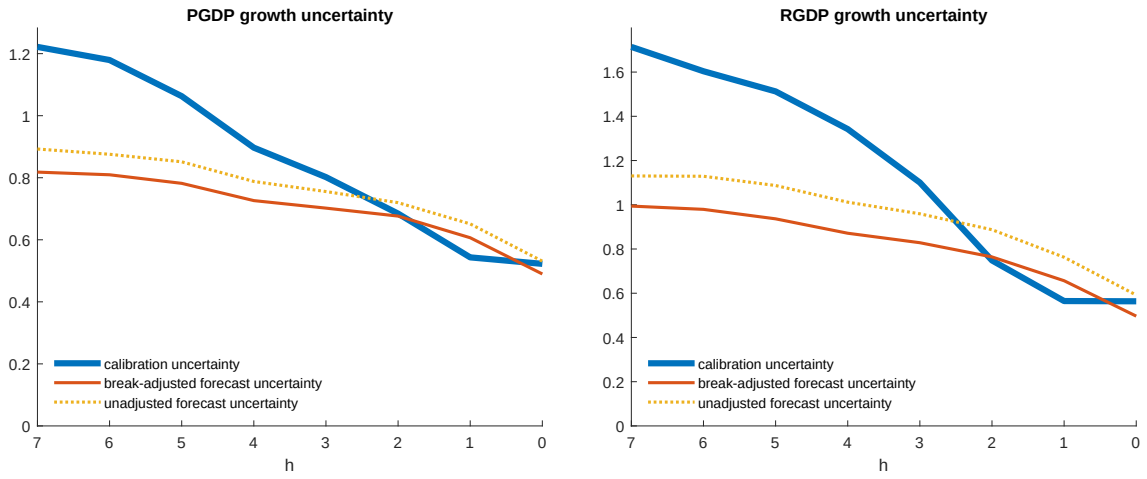

Figure 8: Average forecast uncertainties $\overline{\sigma}_h$ (unadjusted) and $\overline{\tilde{\sigma}}_h$ (break-adjusted) and calibration uncertainty $\sqrt{\overline{S}_h + \overline{D}_h}$ for the sample 1983 to 2024, excluding 2022 for GDP deflator growth (left panel) and 2020 for real GDP growth (right panel)

including the years 2022 (for GDP deflator growth) and 2020 (for real GDP growth) are used, most rounded bin widths have to be 0.5 pp or 1.0 pp larger to approximate the calibration uncertainty, as reported in Table 9 in Appendix D. Based on the non-rounded monotonic bin widths $\widehat{x}_h^*$ in Tables 8 and 9, one can conclude that the bin widths for the largest horizons should be about 2-3 times larger than the bin widths for the shortest horizons.

In analogy to (19), we define the uncertainties based on the rounded and monotonic bin widths $r\left(\widehat{x}_h^*\right)$ by

$$\grave{\sigma}_{h,t}^* = \tilde{\sigma}_{h,t}\left(r\left(\widehat{x}_h^*\right)\right)^{\gamma}$$

$$\overline{\grave{\sigma}}_h^* = \sqrt{\frac{1}{T'-t'+1}\sum_{t=t'}^{T'}\left(\grave{\sigma}_{h,t}^*\right)^2}$$

where $\grave{\sigma}_{h,t}^*$ denotes the aggregate forecast uncertainty and $\overline{\grave{\sigma}}_h^*$ is the corresponding average forecast uncertainty. Figure 9 shows $\overline{\grave{\sigma}}_h^*$ in comparison to the target uncertainty. Deviations from the latter quantity are only due to rounding and imposing monotonicity. Consequently, the deviations are more perceivable if rounding (e.g. for PGDP growth at $h = 3$) or imposing monotonicity (e.g. for RGDP growth at $h = 1$) have a relatively strong effect. Overall, however, the average forecast uncertainty $\overline{\grave{\sigma}}_h^*$ matches the target uncertainty considerably better than $\overline{\sigma}_h$ and $\overline{\tilde{\sigma}}_h$ (see Figure 8).

Figure 10 shows the time series of aggregate forecast uncertainty $\grave{\sigma}_{h,t}^*$ that corresponds to the break-adjusted aggregate uncertainty displayed in Figure 6, i.e. with missing values in 1984q1 and 1985q1, and plotting forecasts *from* (not for) the dates on the x-axis. The current-year uncertainty $\grave{\sigma}_{h,t}^*$ with $h = 0, 1, 2, 3$ and the next-year uncertainty $\grave{\sigma}_{h,t}^*$ with $h = 4, 5, 6, 7$ are further apart than those of the break-adjusted uncertainty displayed in Figure 6. Moreover, the seasonal patterns are more pronounced. Like the break-adjusted uncertainty, $\grave{\sigma}_{h,t}^*$ hardly appears to co-move with the squared forecast errors or the disagreement measures

displayed in Figure 7. Since forecast errors and disagreement are the basis for some common measures of uncertainty (see, for instance, Rossi and Sekhposyan, 2015 and Lahiri and Sheng, 2010), this lack of co-movement seems noteworthy.

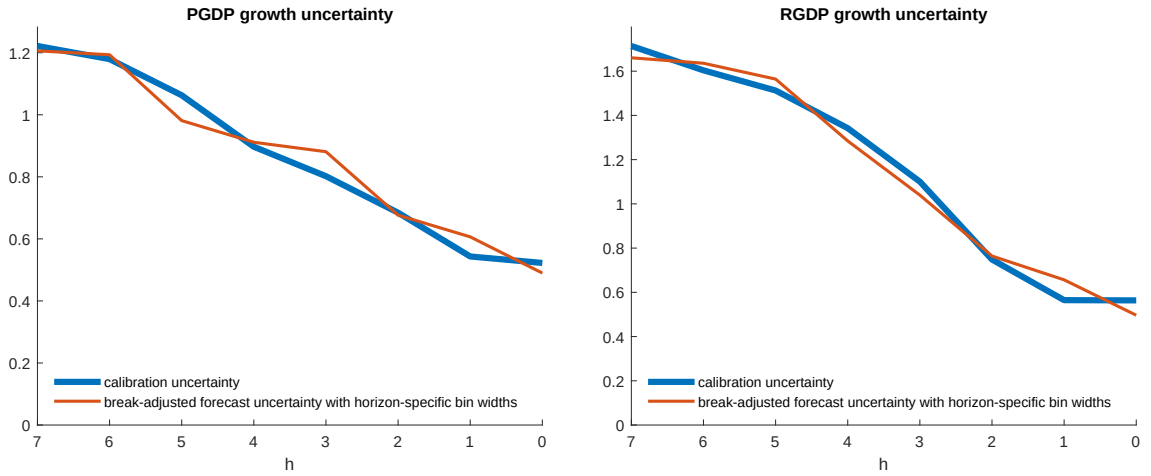

Figure 9: Average forecast uncertainty $\overline{\grave{\sigma}}^*_h$ based on rounded and monotonic bin widths $r\,(\widehat{x}^*_h)$ and calibration uncertainty $\sqrt{\overline{S}_h + \overline{D}_h}$ for the sample 1983 to 2024, excluding 2022 for GDP deflator growth (left panel) and 2020 for real GDP growth (right panel)

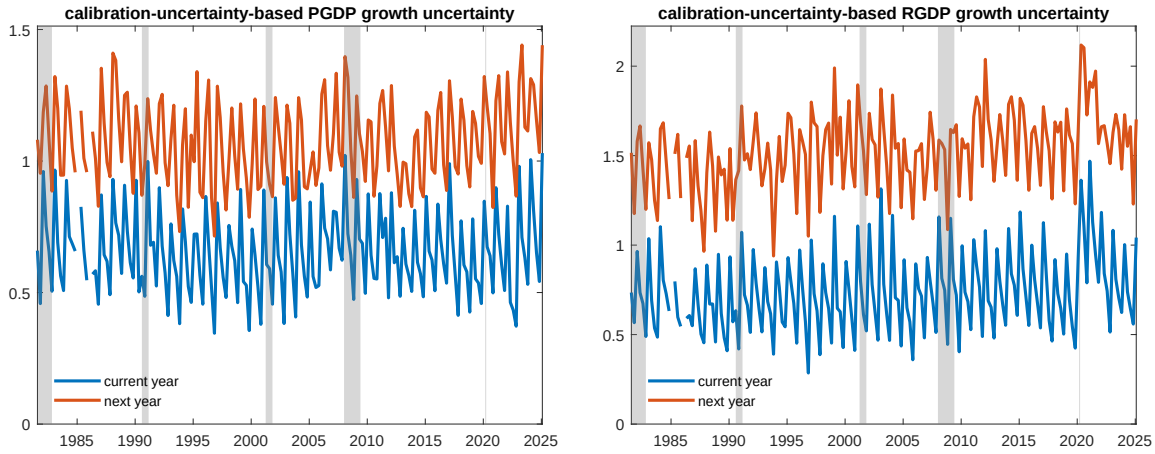

Figure 10: Calibration-uncertainty-based aggregate uncertainty $\grave{\sigma}^*_{h,t}$ of GDP deflator growth (left panel) and of real GDP growth (right panel). Red vertical lines indicate the dates with changes in bin widths. Shaded areas indicate recessions as dated by the NBER.

5 Conclusion

We analyze the aggregate uncertainties for GDP deflator growth and real GDP growth based on the corresponding histogram forecasts of the US SPF. These uncertainties turn out to contain structural breaks caused by changes in the survey design, namely in the widths of the histogram bins. We exploit the structural break in measured aggregate GDP deflator growth uncertainty in 2014 – during a time of virtually constant inflation uncertainty – to

estimate the effect of a bin width change on measured aggregate uncertainty. The estimation result is used to provide break-adjusted aggregate uncertainties for GDP deflator and real GDP growth.

In a second step, we compare the levels of average aggregate uncertainties to their expected values, the calibration uncertainties. Since it turns out that neither the break-adjusted nor the unadjusted uncertainties match the calibration uncertainties, we propose horizon-dependent bin widths that would result in such a match. Rounding to multiples of 0.5 pp and enforcing monotonicity yields bin widths equal to 2.0 pp for inflation and to 2.5 pp for growth for the next-year forecasts made in the first quarter of the year. The bin widths tend to decrease as the forecast horizon shortens, arriving at a bin width of 1.0 pp for both variables, GDP deflator and real GDP growth, for the current-year forecasts in the fourth quarter of the year.

The large effects of changes in the histogram bin widths on measured uncertainty suggest that there are important measurement issues associated with the current practice of eliciting probabilistic forecasts. Therefore, we believe that the effects of survey design need to be studied thoroughly for professional forecasters, just as they are in the case of consumer surveys (see e.g. 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 such investigations.

References

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*(3), 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*(1), 282–303. Annals Issue: Subjective Expectations & Probabilities in Economics.

Andrade, P., E. Ghysels, and J. Idier (2012). Tails of Inflation Forecasts and Tales of Monetary Policy. Working papers 407, Banque de France.

Bai, J. (1997). Estimating multiple breaks one at a time. *Econometric Theory 13*(3), 315–352.

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica 66*(1), 47–78.

Bai, J. and P. Perron (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics 18*(1), 1–22.

Baker, S. R., N. Bloom, and S. J. Davis (2016, 07). Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics 131*(4), 1593–1636.

Becker, C., P. Duersch, and T. Eife (2023, January). Measuring Inflation Expectations: How the Response Scale Shapes Density Forecasts. Working Papers 0723, University of Heidelberg, Department of Economics.

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

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica 77*(3), 623–685.

Broer, T. and A. N. Kohlhas (2024). Forecaster (mis-)behavior. *The Review of Economics and Statistics 106*(5), 1334–1351.

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? Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3894581, forthcoming in the Journal of Econometric Literature.

Clark, T., G. Ganics, and E. Mertens (2022). What is the predictive value of spf point and density forecasts? Working Papers 22-37, Federal Reserve Bank of Cleveland.

Clements, M. P. (2014). Forecast uncertainty—ex ante and ex post: U.s. inflation and output growth. *Journal of Business & Economic Statistics 32*(2), 206–216.

Clements, M. P. (2018). Are macroeconomic density forecasts informative? *International Journal of Forecasting 34*(2), 181–198.

Clements, M. P., R. W. Rich, and J. S. Tracy (2023). Surveys of Professionals. In R. Bachmann, G. Topa, and W. van der Klaauw (Eds.), *Handbook of Economic Expectations*, pp. 71–106. Academic Press.

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

de Bruin, W. B., 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.

Diebold, F. X., A. S. Tay, and K. F. Wallis (1999). Evaluating density forecasts of inflation: the survey of professional forecasters. In R. Engle and H. White (Eds.), *Festschrift in Honor of C.W.J. Granger*, pp. 76–90. Oxford: Oxford University Press.

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*(1), 30–41.

Ganics, G., B. Rossi, and T. Sekhposyan (2024). From fixed-event to fixed-horizon density forecasts: Obtaining measures of multihorizon uncertainty from survey density forecasts. *Journal of Money, Credit and Banking 56*(7), 1675–1704.

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

Glas, A. and M. Hartmann (2022). Uncertainty measures from partially rounded probabilistic forecast surveys. *Quantitative Economics 13*(3), 979–1022.

Goldfayn-Frank, O., P. Kieren, and S. T. Trautmann (2025). A choice-based approach to the measurement of inflation expectations. *Journal of Monetary Economics*, 103882.

Jurado, K., S. C. Ludvigson, and S. Ng (2015, March). Measuring uncertainty. *American Economic Review 105*(3), 1177–1216.

Knüppel, M. and F. Krüger (2022). Forecast uncertainty, disagreement, and the linear pool. *Journal of Applied Econometrics 37*(1), 23–41.

Krüger, F. and L. Pavlova (2024). Quantifying subjective uncertainty in survey expectations. *International Journal of Forecasting 40*(2), 796–810.

Lahiri, K. and X. Sheng (2010). Measuring forecast uncertainty by disagreement: The missing link. *Journal of Applied Econometrics 25*(4), 514–538.

Lahiri, K. and W. Wang (2020). Estimating macroeconomic uncertainty and discord using info-metrics. In M. Chen, J. M. Dunn, A. Golan, and A. Ullah (Eds.), *Advances in Info-Metrics: Information and Information Processing across Disciplines*, Chapter 11, pp. 290–324. New York, NY: Oxford University Press.

Lichtendahl, K. C., Y. Grushka-Cockayne, and R. L. Winkler (2013). Is it better to average probabilities or quantiles? *Management Science 59*(7), 1594–1611.

Liu, Y. and X. S. Sheng (2019). The measurement and transmission of macroeconomic uncertainty: Evidence from the u.s. and bric countries. *International Journal of Forecasting 35*(3), 967–979. Forecasting issues in developing economies.

Ludvigson, S. C., S. Ma, and S. Ng (2021, October). Uncertainty and business cycles: Exogenous impulse or endogenous response? *American Economic Journal: Macroeconomics 13*(4), 369–410.

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*(5), 1464–1476.

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*(1), 200–207.

Rossi, B. and T. Sekhposyan (2015, May). Macroeconomic uncertainty indices based on nowcast and forecast error distributions. *American Economic Review 105*(5), 650–55.

Wallis, K. F. (2005). Combining density and interval forecasts: A modest proposal*. *Oxford Bulletin of Economics and Statistics 67*(s1), 983–994.

A Appendix: 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. Let B be the length of the base of the triangle, i.e. $B = w$. 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, so that $B = 2w$. The corresponding triangular distribution has the mean $\mu = A + B/2$ and the variance $\sigma^2 = B^2/24$.

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. The corresponding triangular distribution has the mean $\mu = A - v + B/2$ and the variance $\sigma^2 = B^2/24$. One therefore needs to calculate the length of the base of the triangle, B, as done below.

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 the right endpoint b

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Case 2.3.2 if $q_v = \frac{1}{2}$, then we can simply pin $a = A - v$ and $b = A + w$

Case 3 The forecasters uses three or more bins. In this case, we assume a normal distribution and minimize the function

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

where K is the number of bins, A_k denotes the upper bound of the k-th interval, $\Phi(A_k, \mu, \sigma)$ is the value of the CDF of the normal distribution with mean μ and standard deviation σ at A_k, and q_k denotes the probability assigned to the k-th interval by the forecaster. We have $\Phi(A_0, \mu, \sigma) = 0$ and $\Phi(A_1, \mu, \sigma) = 1$ because of $A_0 = -\infty$ and $A_K = \infty$.

In rare cases, the fitted standard deviation of a normal distribution can be marginally smaller than the standard deviation of a one-bin triangular distribution. This happens, for instance, if the probabilities 0.1, 99.8, and 0.1 are assigned to three adjacent bins with equal width. In these rare cases, we set the standard deviation to the value corresponding to a one-bin triangular distribution applied to the narrowest bin.

B Appendix: Seasonally-adjusted inflation uncertainty series

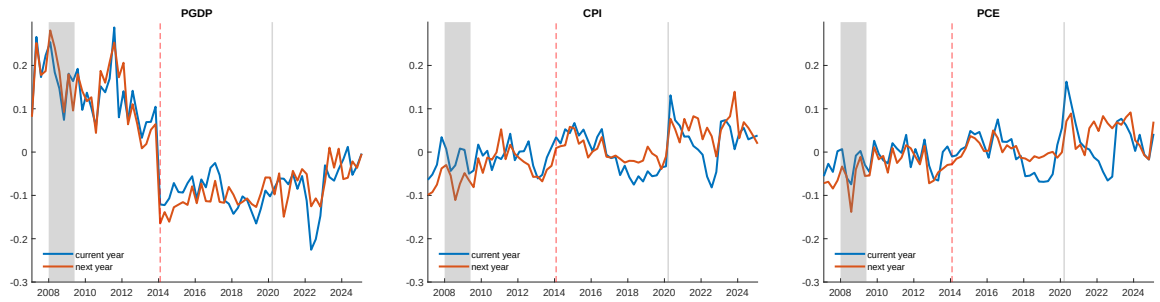

Figure 11: Seasonally-adjusted residual series of GDP deflator growth (left panel), core CPI inflation (middle panel), and core PCE inflation (right panel) uncertainty. The red vertical line indicates date of change in bin widths for GDP deflator growth. Shaded areas indicate recessions as dated by the NBER.

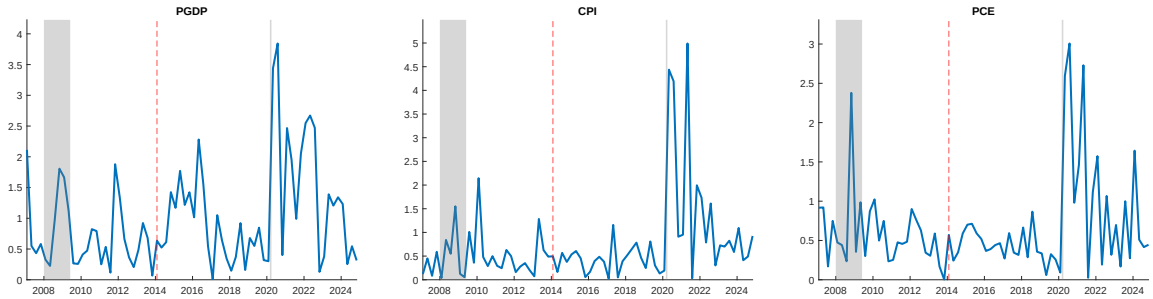

Figure 12: Absolute values of residuals of AR(1)-processes for seasonally-adjusted GDP deflator growth (left panel), core CPI inflation (middle panel), and core PCE inflation (right panel). The red vertical line indicates date of change in bin widths for GDP deflator growth. Shaded areas indicate recessions as dated by the NBER.

C Estimates of individual elasticities

We estimate the individual values of γ_n by considering each forecaster in the sample from 2010 to 2019 who made at least one forecast before 2014 and at least one forecast since 2014. One could attain more precise individual estimates by requiring more observations per subsample, but this would come at the cost of excluding more forecasters. As for the estimation of the aggregate elasticity γ, we impose the restriction that γ_n does not depend on the forecast horizon. For each forecaster, we employ the specification of regression equation (8) with that forecaster's uncertainty $\sigma^{\pi}_{n,j|t,i}$ as the dependent variable. Hence, we estimate

$$\ln\left(\sigma^{\pi}_{n;j|t,i}\right) = c_{n;i,j} + \gamma_n \ln\left(x^{\pi}_{t,i}\right) + \varepsilon_{n;t,i,j} \tag{22}$$

for each n, where $x^{\pi}_{t,i}$ is the same as in equation (8).

Note that the estimate of the aggregate elasticity γ relies on a sample containing all forecasters who made at least one forecast in the sample 2010 to 2019. Thus, the number of forecasters entering the estimation of γ is larger than in the sample used to estimate the γ_n's, although we consider the largest possible number of forecasters for the latter case.

Our sample to estimate the individual γ_n's contains 46 forecasters. The histogram of these estimates, displayed in Figure 13, has a hump-shaped form. Thus, there is no evidence of bimodality, suggesting the absence of distinct groups of forecasters with $\gamma_n \approx 0$ and $\gamma_n \approx 1$, respectively. There are estimates below 0 and above 1, which are likely due to parameter estimation error. The average across all estimated γ_n's equals 0.64, which is close to the value of 0.56 estimated for the aggregate elasticity γ. If we set the outliers being smaller than 0 to 0, and the outliers being larger than 1 to 1, the average across the γ_n's equals 0.620. Note that the relation between the γ_n's and the aggregate elasticity γ depends on the correlation between the individual uncertainty forecasts $\sigma^{\pi}_{n;j|t,i}$ and the individual elasticities γ_n. If, for instance, forecasters with larger elasticities also tend to report higher uncertainties, γ will exceed the average across the γ_n's.

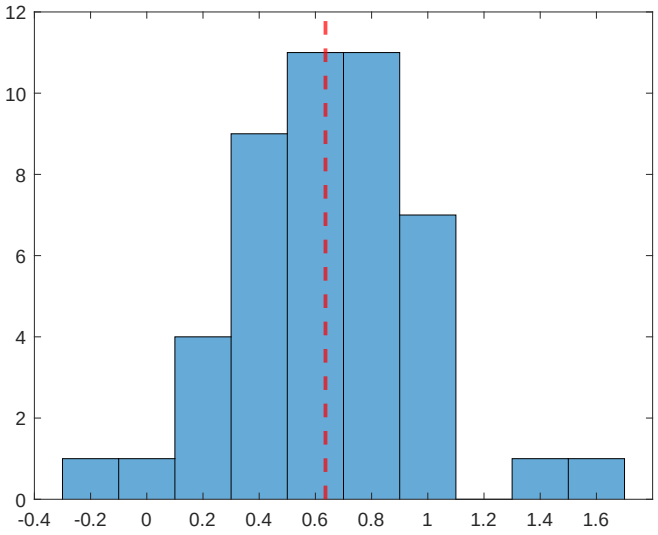

Figure 13: Histogram of estimates of γ_n based on the sample 2010 to 2019. The dashed red line indicates the average across the estimates.

D Appendix: Full-sample results for average forecast uncertainty

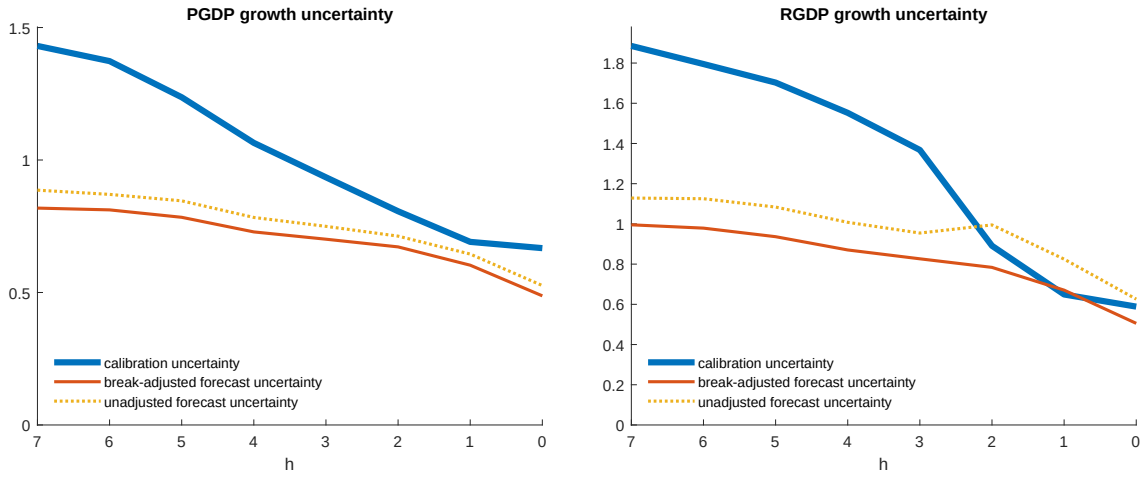

Figure 14: Average forecast uncertainties $\overline{\sigma}_h$ (unadjusted) and $\overline{\tilde{\sigma}}_h$ (break-adjusted) and calibration uncertainty $\sqrt{\overline{S}_h + \overline{D}_h}$ for GDP deflator growth (left panel) and real GDP growth (right panel) for the full sample 1983 to 2024

Table 9: Calibration-uncertainty-based bin widths, full sample

	PGDP growth				RGDP growth			
	x_h^*	$\widehat{x}_h^*$	$r\left(x_h^*\right)$	$r\left(\widehat{x}_h^*\right)$	x_h^*	$\widehat{x}_h^*$	$r\left(x_h^*\right)$	$r\left(\widehat{x}_h^*\right)$
$h=7$	2.71	2.71	2.50	2.50	3.13	3.13	3.00	3.00
$h=6$	2.56	2.56	2.50	2.50	2.95	2.95	3.00	3.00
$h=5$	2.26	2.26	2.50	2.50	2.91	2.91	3.00	3.00
$h=4$	1.97	1.97	2.00	2.00	2.81	2.81	3.00	3.00
$h=3$	1.67	1.67	1.50	1.50	2.46	2.46	2.50	2.50
$h=2$	1.39	1.47	1.50	1.50	1.26	1.26	1.50	1.50
$h=1$	1.28	1.47	1.50	1.50	0.94	1.13	1.00	1.00
$h=0$	1.75	1.47	2.00	1.50	1.31	1.13	1.50	1.00

Note: US SPF bin widths (in pp) making the average forecast uncertainty (approximately) equal to the calibration uncertainty for the sample 1983 to 2024. x_h^* leads to equality, $\widehat{x}_h^*$ enforces monotonicity with respect to h. $r(\bullet)$ rounds both types of bin widths to multiples of 0.5.

Download ZEW Discussion Papers:

https://www.zew.de/en/publications/zew-discussion-papers

or see:

https://www.ssrn.com/link/ZEW-Ctr-Euro-Econ-Research.html
https://ideas.repec.org/s/zbw/zewdip.html

IMPRINT

ZEW – Leibniz-Zentrum für Europäische Wirtschaftsforschung GmbH Mannheim
ZEW – Leibniz Centre for European Economic Research

L 7,1 · 68161 Mannheim · Germany
Phone +49 621 1235-01
info@zew.de · zew.de