Federal Reserve Bank of Cleveland Working Paper Series

Surveys of Professionals

Michael P. Clements, Robert W. Rich, and Joseph S. Tracy

Working Paper No. 22-13

May 2022

Suggested citation: Clements, Michael P., Robert W. Rich, and Joseph S. Tracy. 2022. "Surveys of Professionals." Working Paper No. 22-13. Federal Reserve Bank of Cleveland. https://doi.org/10.26509/frbc-wp-202213.

Federal Reserve Bank of Cleveland Working Paper Series
ISSN: 2573-7953

Surveys of Professionals

Michael P. Clements
Robert Rich
Joseph Tracy*

April 26, 2022

Forthcoming in *Handbook of Economic Expectations*, Elsevier.

Abstract: This chapter provides an overview of surveys of professional forecasters, with a focus on the U.S. Survey of Professional Forecasters and the European Central Bank Survey of Professional Forecasters. A distinguishing feature of these surveys is that they collect point and density forecasts and make the data publicly available. We discuss their structure, issues involved in using the data, and the construction of measures such as disagreement and uncertainty at the aggregate and individual levels. Our review also summarizes the findings of studies exploring issues such as the alignment of point forecasts with measures of central tendency from associated density forecasts, the coverage of density forecasts, the rounding of point and density forecasts, comparisons of forecast accuracy across respondents, and heterogeneity in forecast behavior and the persistence of these differential features. We conclude with some observations for future work.

Keywords: economic forecasting; survey expectations; data collection and modeling; quantitative methods
JEL codes: C53; E37; D84

* Clements: ICMA Centre, Henley School of Business, University of Reading; Rich: Federal Reserve Bank of Cleveland; Tracy: Federal Reserve Bank of Dallas. The views expressed in the paper are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Cleveland, the Federal Reserve Bank of Dallas, or the Federal Reserve System. Address for correspondence: Research Department, P.O. Box 6387, Federal Reserve Bank of Cleveland, Cleveland, OH 44101-1387. Phone: 216-579-2928. Email: Robert.rich@clev.frb.org
We are grateful to the Handbook editors as well as Dean Croushore and Geoff Kenny for helpful comments. We would also like to thank Amber Sherman for her assistance with the Appendix Table that accompanies this chapter.

I. Surveys of Professional Forecasters

a. Interest and Attractiveness of Eliciting Professional Forecasters' Expectations

Expectations play a critical role in agents' decision-making processes and are a key determinant of many economic and financial variables. This chapter considers professional forecasters, who are typically assumed to possess extensive and largely homogeneous information sets, as well as sophisticated and comparable processing capabilities. Surveys of households and firms are covered in Chapter 1: "Household Surveys and Probabilistic Questions," and Chapter 2: "Firm Surveys," respectively. Chapter 18: "Inference on Probabilistic Surveys in Macroeconomics with an Application to the Evolution of Uncertainty in the Survey of Professional Forecasters during the COVID Pandemic" focuses on conducting inference and complements this chapter.

Professional forecasters are an attractive source of expectations data for researchers and policymakers. For example, Keane and Runkle (1990) argue that testing for rationality might be more appropriately applied to professional forecasters rather than to other respondents. Professional forecasters should be better informed and better able to respond to technical questions, such as the formulation and reporting of "probabilistic beliefs" using histograms. In addition, professional forecasters mitigate the need to control for demographics and other individual-specific characteristics. Consequently, surveys of professional forecasters provide data that are useful in monitoring movements in expectations and evaluating their influence. Surveys of professionals can also be used to monitor central bank credibility (see, e.g., Beechey, Johannsen, and Levin, 2011, and Dovern and Kenny, 2020), evaluate their influence on private-sector expectations (e.g., Carroll, 2003), and incorporate direct measures of expectations into economic models (see, e.g., Smets, Warne, and Wouters, 2014).

This chapter examines surveys of professional forecasters, with a focus on the U.S. Survey of Professional Forecasters and the European Central Bank Survey of Professional Forecasters. A distinguishing feature of these surveys is that they collect point and density forecasts and make the data publicly available. We discuss the history and design of the surveys, issues involved in using the data, and the construction of measures such as disagreement and uncertainty. An Appendix Table is available that provides background information on a large set of primary surveys of professional forecasters. The chapter also summarizes the findings of studies that have used surveys of professional forecasters to explore a wide range of issues. These include heterogeneity and persistence in individual forecast behavior, comparisons of the properties of point and density forecasts, and the reliability of using disagreement as a proxy for uncertainty.

b. Surveys of Professional Forecasters

U.S. Survey of Professional Forecasters

The U.S. Survey of Professional Forecasters (US-SPF) is a survey of U.S. macro-variables. The survey began in the fourth quarter of 1968 and was initially conducted by the American Statistical Association and the National Bureau of Economic Research, but from 1990 to the present day, it has been run by the Federal Reserve Bank of Philadelphia. See Zarnowitz (1969) on the original objectives of the survey, and Croushore (1993) and Croushore and Stark (2019) on the revival of the survey by the Federal Reserve Bank of Philadelphia. The survey is conducted each quarter after the release of the advance report of the national income and product accounts (NIPA) from the Bureau of Economic Analysis (BEA). Since 1990:Q2, the deadlines for responses have been around the middle of each quarter, before the BEA's second report. The average number of panelists has been around 40 since the Philadelphia Fed has overseen the survey. The respondents are anonymous but have identifiers allowing the forecasts of a given individual to be identified across surveys. There is a quarterly report summarizing the results of the survey and the data are available on the Federal Reserve Bank of Philadelphia's website.

European Central Bank Survey of Professional Forecasters

The European Central Bank Survey of Professional Forecasters (ECB-SPF) provides a survey of forecasts for the euro area. The survey began in January 1999 and draws respondents from both financial and non-financial institutions, with most, but not all, located in the euro area. The principal aim of the survey is to elicit expectations about inflation, real GDP growth, and unemployment, although the questionnaire also contains a non-compulsory section asking participants for their expectations of additional variables and to provide qualitative comments that inform their quantitative forecasts.[1] The survey is typically fielded in February, May, August, and November, with a little under 60 respondents on average per survey.[2] While there is a listing of the respondents' names and institutions, individual responses are anonymous and tracked across surveys by an assigned identification number. Similar to the US-SPF, there is a quarterly report summarizing

[1] The additional expectations pertain to variables such as wage growth, the price of oil, and the exchange rate. See Meyler (2020).

[2] The actual average of 58 panelists per survey is based on the history of the ECB-SPF, although there has been some decline in participation over time.

the results of the survey and the data are available on the ECB's website. For additional details about the ECB-SPF, see Garcia (2003) and Bowles et al. (2007).

Generally, we do not know how the respondents make their forecasts in any specific instance. An ad hoc US-SPF survey by Stark (2013) found the majority "use a mathematical/computer model plus subjective adjustments to that model in reporting their projections." No information is available in the US-SPF on forecaster characteristics, except whether the company is in the financial or non-financial services sector (if known). Special surveys conducted by the ECB in 2008, 2013, and 2018 indicate that most participants report judgment-based forecasts, although the number of model-based forecasts appears to be increasing over time. The ECB-SPF does not provide any information on forecaster characteristics. The respondent anonymity of each survey may discourage respondents from strategic reporting of forecasts.

c. Nature of Survey Expectations and Concepts

The US-SPF and ECB-SPF elicit point forecasts for specified variables and horizons. Point forecasts provide direct measures of expectations and avoid the challenges of model-based measures (for example, the specification of regression equations and information sets). In addition, even among professional forecasters there is disagreement as evidenced by the dispersion in predictions.

There are, however, limitations to point forecast data. One is their interpretation: a point forecast may represent a mean, median, or mode, or some other construct from the respondent's subjective probability distribution. Because surveys do not provide any guidance on loss functions, it is unclear for any given respondent what the point forecast represents and the degree to which the possible representations may vary across respondents.[3] We return to this issue in Section IV.

The US-SPF and ECB-SPF also ask forecasters to report density forecasts using histograms with a set of intervals provided in the survey instrument. Density forecasts provide a basis for constructing measures of uncertainty, and the shape of a density forecast can be informative about the respondent's balance of risks. The extent of skewness displayed by a density forecast reflects the respondent's assessment of upside or downside risks. Through the ability to convey uncertainty and balance of risks, density forecasts provide survey participants with a medium in which to describe

[3] If the respondent chooses to minimize a squared-error loss function, then the conditional mean will be reported as the forecast. The three special ECB surveys in 2008, 2013, and 2018 also asked participants if their reported point forecasts refer to the mean, median, or mode of their subjective probability distributions. While most participants initially indicated that the point forecasts refer to the mean, the median and the mode are now being reported more often. Comparable information for the US-SPF is not available.

their outlook in more detail and allow for a more thorough analysis of forecast properties than can be carried out using point forecasts alone. Nevertheless, it is worth noting that histogram-based density forecasts also present respondents and researchers with several challenges, which we discuss below.

II. Point and Density Forecasts: Data Features, Measures, and Properties

a. Background

Respondent Participation Patterns/Survey Design

The US-SPF and ECB-SPF involve panels of forecasters whose composition changes as some respondents leave and new respondents are added. The panels are unbalanced due both to entry and exit and to the fact that there are occasions when respondents do not provide responses to all or part of the survey questionnaire. Respondents' numerical identifiers allow their forecasts to be tracked over time.[4] This supports individual-level analyses of forecast behavior and allows comparisons between forecasters, such as whether some respondents are relatively more accurate or tend to be more contrarian or more uncertain compared to others.

The coverage of many surveys has expanded over time, with questions added to elicit expectations about new variables, concepts (such as the natural rate of unemployment), as well as density forecasts of some variables and longer-horizon forecasts.[5] Nevertheless, forecasts may not be available at the specific times and horizons required by the survey user. An example is when the analyst requires a series of fixed-horizon forecasts, but the forecast data are instead fixed-event, such as with year-on-year calendar growth rates. As discussed below, various techniques have been suggested to deal with these limitations.

The vague phrasing of survey questionnaires may complicate the task of the survey respondents and users in various ways. First, respondents forecasting target variables subject to revision are not provided guidance about the specific vintage to consider. In the case of GDP, for example, it is not clear whether an individual should provide his or her "best" estimate of the first release (i.e., advance estimate), the first quarterly revised estimate, or the final estimate for the target variable. Only a few

[4] It is not always clear whether the identifiers are associated with the individuals or the companies that employ them, and what happens to the identifier when an employee changes firms. In its documentation, the US-SPF states that the identifier either moves with the individual or stays with the company depending on whether "a forecast seems associated more with the firm than the individual."

[5] For example, the US-SPF began asking for forecasts of the natural rate of unemployment in 1996:Q3 (for the third-quarter surveys), density forecasts for the unemployment rate in 2009:Q2, and forecasts for 10-year annual-average real GDP growth and productivity growth in 1992:Q1 (for first-quarter surveys).

variables are not subject to revision, while many are heavily revised.[6] Second, as already noted, it is unclear how the respondents' point forecasts relate to their (assumed) subjective distributions of the target variables. We discuss this in Section IV.

Less obviously, the question of what the forecasts represent also applies to the density forecasts. For example, an analyst might evaluate a respondent's density forecasts using a popular scoring rule, such as the quadratic probability score (QPS: Brier (1950)) or the ranked probability score (RPS: Epstein (1969)), and find in favor of an alternative set of forecasts, such as benchmark forecasts. But the respondent might be driven by motives other than maximizing QPS or RPS, such as ensuring that the realized outcomes fall well within the likely range of outcomes implied by their probability assessments.

Target Variables and Forecast Horizons

The surveys typically include a mixture of fixed- and rolling-event (or fixed-horizon) forecasts. For example, the US-SPF asks for density forecasts of the annual rate of GDP deflator inflation in the year of the survey relative to the previous year and of the next year relative to the current year, that is, of the percentage rate of change in the annual GDP deflator between years. This means that there are eight "fixed-event" histogram forecasts of annual inflation in 2016, compared to 2015. The first was made in response to the 2015:Q1 survey, with a horizon of nearly two years. The second was made in 2015:Q2, and so on, down to the last one made in 2016:Q4 (with a horizon of just under a quarter). An *annual* series of fixed-horizon histograms of approximately one year can be constructed using the first-quarter survey responses. However, some analyses require quarterly rather than annual series.

A simple way of constructing an approximate series of quarterly year-ahead forecasts has been proposed by D'Amico and Orphanides (2008). They suggest using the first-quarter current-year forecast, the second-quarter current-year forecast, the third-quarter next-year forecast, and the fourth-quarter next-year forecast. This gives a quarterly-frequency series of forecasts, although from the first to the fourth quarters the actual horizons are 4, 3, 6, and 5, and the target moves from the current to the next year's growth rate between the second and third quarters. An alternative solution is to take a weighted average of the current and next year's forecasts, where the weights vary with the quarter of the year of the survey, reflecting the distances of the forecasts from the desired forecast

[6] Reviews of data revisions and real-time analysis are provided by Croushore (2006, 2011a, 2011b) and Clements and Galvão (2019).

horizon, as suggested by D'Amico and Orphanides (2014). A recent contribution by Ganics, Rossi, and Sekhposyan (2020) suggests weighting together the fixed-event density forecasts to obtain uniformity of the probability integral transform of the combined (fixed-horizon) density. The aim is to obtain a correctly calibrated fixed-horizon density forecast.

In terms of point forecasts, the US-SPF typically provides forecasts of the current quarter (i.e., of the quarter in which the survey is held) and of each of the next four quarters, as well as the current year and the following year (including density forecasts). If we denote the survey quarter by *t*, the forecast horizons are $h = 0, 1, 2, 3, 4$, where $h = 4$ indicates a forecast of the same quarter of the year (as the quarter of the survey) in the following year. Hence, we have quarterly series of rolling-event forecasts, or fixed-horizon forecasts, for $h = 0, 1, 2, 3, 4$. But when the data are linked to earlier surveys, forecasts with a fixed-event structure also arise.

The nature of the US-SPF point forecasts is as described in Table 1, where $y_{t+h\,|\,t}$ denotes a forecast of y_{t+h} made at time *t*.

Table 1: The Fixed-Event/Fixed-Horizon Nature of the US-SPF Point forecasts

Survey/Target	y_{t-2}	y_{t-1}	y_t	y_{t+1}	y_{t+2}	y_{t+3}	y_{t+4}	y_{t+5}	
$\vdots$									
$t-2$	$y_{t-2\,|\,t-2}$	$y_{t-1\,|\,t-2}$	$y_{t\,|\,t-2}$	$y_{t+1\,|\,t-2}$	$y_{t+2\,|\,t-2}$				
$t-1$		$y_{t-1\,|\,t-1}$	$y_{t\,|\,t-1}$	$y_{t+1\,|\,t-1}$	$y_{t+2\,|\,t-1}$	$y_{t+3\,|\,t-1}$			
t			$y_{t\,|\,t}$	$y_{t+1\,|\,t}$	$y_{t+2\,|\,t}$	$y_{t+3\,|\,t}$	$y_{t+4\,|\,t}$		
$t+1$				$y_{t+1\,|\,t+1}$	$y_{t+2\,|\,t+1}$	$y_{t+3\,|\,t+1}$	$y_{t+4\,|\,t+1}$	$y_{t+5\,|\,t+1}$	
$t+2$					$y_{t+2\,|\,t+2}$	$y_{t+3\,|\,t+2}$	$y_{t+4\,|\,t+2}$	$y_{t+5\,|\,t+2}$	$\ddots$
$t+3$						$y_{t+3\,|\,t+3}$	$y_{t+4\,|\,t+3}$	$y_{t+5\,|\,t+3}$	$\ddots$
$t+4$							$y_{t+4\,|\,t+4}$	$y_{t+5\,|\,t+4}$	$\ddots$
$t+5$								$y_{t+5\,|\,t+5}$	$\ddots$
$\vdots$									$\ddots$

The five diagonals give the fixed-horizon forecasts, with the bottom diagonal containing the $h = 0$ forecasts, the one above the $h = 1$ forecasts, and so on up to the top diagonal containing the longest-horizon forecasts. The fixed-event forecasts are given by the columns. For example, reading the column headed y_{t+3}, from bottom to top gives the $h = 0$ to $h = 4$ forecasts of y_{t+3}.

In the context of obtaining fixed-horizon point forecasts from fixed-event point forecasts, Knüppel and Vladu (2016) cite a large number of papers that have used ad hoc weights. They provide a general framework for obtaining optimal approximations—in the sense of minimizing the mean-squared error—to fixed-horizon forecasts when the requisite forecasts are not provided. They illustrate by calculating one-year-ahead inflation and growth forecasts from Consensus Economics quarterly forecasts of annual inflation and growth in the current year and the next year.

The ECB-SPF provides fixed- and rolling-event forecasts at short-, medium- and longer-term horizons. In terms of fixed-event horizons, the survey asks for forecasts of the current calendar year and the next two calendar years. For inflation and growth, the forecasts for the current calendar year and the next calendar year parallel those in the US-SPF. For unemployment, the current calendar year and the next calendar year refer to the average of monthly unemployment rates in the current year and the subsequent year of the survey, respectively. The remaining calendar year horizon for inflation, growth, and unemployment follows analogously with the appropriate shifting forward of reference years.

Compared to the US-SPF, the ECB-SPF provides four additional fixed-event forecasts of the target variables due to the horizon extending out to a maximum of three years. There are two other important differences between the surveys. First, the ECB-SPF provides both point and density forecasts for the calendar year horizons allowing for matched series across all 12 fixed-event forecasts for a particular calendar year. Second, the ECB-SPF also provides matched point and density rolling-event forecasts for inflation, growth, and unemployment at the one-year-ahead and the one-year/one-year-forward horizons. This structure yields *quarterly* series of fixed horizon forecasts that offer a three-fold increase in the number of observations compared to the US-SPF.[7]

b. Point Forecasts

Measures at the Aggregate and Individual Levels

While surveys of professional forecasters record expectations data at the individual level, the fielding agencies typically report aggregate responses, such as the mean or median of the individual responses. These measures may suffice for empirical studies that only require observations on a "consensus" forecast. On other occasions, the individual responses are required, as when testing the

[7] The ECB-SPF also provides matched longer-term point and density forecasts that correspond to a 4- to 5-year-ahead forecast horizon.

implications of expectations models for features of forecast behavior such as disagreement. For example, disagreement cannot arise under rational expectations when the structure of the economy is known to all forecasters and all information is common. Disagreement, however, may arise for various reasons, including information rigidities (see, e.g., Coibion and Gorodnichenko (2012, 2015b)); differential interpretation of public information (see, e.g., Manzan (2011)); some forecasters receiving superior signals; differential rates of learning in non-stationary environments; and differential loss functions (see, e.g., Capistrán and Timmermann (2009)).

The interest in forecaster disagreement is not limited to its relevance for expectations models. As we discuss later in this section, it remains common practice in empirical studies to use disagreement as a proxy for uncertainty when direct measures for uncertainty are not available. In addition, differences in agents' expectations have been cited as an important channel through which monetary policy can affect real activity [Woodford (2003), Mankiw and Reis (2002), and Lorenzoni (2009)] and a key factor influencing the effect of public information signals (Morris and Shin (2002), and Amador and Weill (2010)). The role of heterogeneous beliefs has also been advanced in terms of explaining the evolutions that led to the global financial crisis [Sims (2009), Tomura (2013), Favara and Song (2014), Tian and Yan (2009), and Geanakoplos (2010)].

Disagreement – Measures of Dispersion

Disagreement is measured by the distance between (point) forecasts, where different choices of the distance metric give rise to alternative disagreement measures. A common measure of disagreement is the cross-sectional standard deviation of the point forecasts (using the squared distance norm from the mean of the point forecasts):

$$s_{t+h\,|\,t} = \sqrt{\frac{1}{N_t - 1}\sum_{i=1}^{N_t}\left(y_{i,t+h\,|\,t} - \overline{y}_{t+h\,|\,t}\right)^2} \tag{1}$$

where $y_{i,t+h\,|\,t}$ denotes the *h*-step-ahead point prediction of forecaster *i* at time *t* and

$\overline{y}_{t+h\,|\,t} = N_t^{-1}\sum_{i=1}^{N_t} y_{i,t+h\,|\,t}$ denotes the mean (or ""consensus") forecast.

Disagreement can also be measured using the average absolute value distance from the consensus forecast or by range-based statistics that are more robust to outliers. With regard to the latter, Abel et al. (2016), Glas and Hartman (2016), and Lahiri and Sheng (2010) use the interquartile range (IQR) of the point forecasts:

$$y_{t+h\,|\,t}^{IQR} = y_{t+h\,|\,t}^{0.75} - y_{t+h\,|\,t}^{0.25} \tag{2}$$

where $y_{t+h\,|\,t}^{0.75}$ and $y_{t+h\,|\,t}^{0.25}$ denote the 75th and 25th percentiles of the ordered array of point forecasts. Giordani and Söderlind (2003) and Boero, Smith, and Wallis (2015) use the quasi-standard deviation – half the distance between the 84th and 16th percentiles of the cross-sectional distribution– which equals one standard deviation if the point forecasts are normally distributed.

Almost all survey-based measures of disagreement are constructed at the aggregate level. Rich and Tracy (2021a), however, have proposed univariate disagreement measures for individual ECB-SPF respondents. Specifically, they define the average absolute point disagreement (AAPD) measure for the *i*th respondent as:

$$AAPD_{i,t+h\,|\,t} = \frac{1}{N_t - 1} \sum_{j \neq i} \left| y_{j,t+h\,|\,t} - y_{i,t+h\,|\,t} \right| \tag{3}$$

The AAPD measure uses an individual forecaster's point prediction as the reference point for the calculation.[8] In particular, the extent of conformity in (3) is based on pairwise comparisons between a respondent's point forecast and those of all the other respondents.

Clements (2021b) constructs multivariate disagreement measures for individual US-SPF respondents. The measure accounts for beliefs about potential interdependencies in the set of variables being forecast. Following Banternghansa and McCracken (2009), the cross-sectional forecast covariance matrix is:

$$\mathbf{S}_{t+h\,|\,t} = N_t^{-1} \sum_{i=1}^{N_t} \left(\mathbf{y}_{i,t+h\,|\,t} - \overline{\mathbf{y}}_{t+h\,|\,t} \right) \left(\mathbf{y}_{i,t+h\,|\,t} - \overline{\mathbf{y}}_{t+h\,|\,t} \right)' \tag{4}$$

where $\mathbf{y}_{i,t+h\,|\,t}$ is the vector of forecasts made by respondent *i* at time *t* for the target variables $\mathbf{y}_{t+h}$, N_t is the number of participants at time *t*, and $\overline{\mathbf{y}}_{t+h\,|\,t} = N_t^{-1} \sum_{i=1}^{N_t} \mathbf{y}_{i,t+h\,|\,t}$ is the cross-sectional average.

[8] As discussed in Rich and Tracy (2021a), the squared distance norm can be used as an alternative in (3) to construct a measure of individual disagreement.

For respondent i, the multivariate disagreement measure is given by:

$$D_{i,t+h\mid t} = \sqrt{\left(\mathbf{y}_{i,t+h\mid t} - \overline{\mathbf{y}}_{t+h\mid t}\right)' \mathbf{S}_{t+h\mid t}^{-1} \left(\mathbf{y}_{i,t+h\mid t} - \overline{\mathbf{y}}_{t+h\mid t}\right)} \tag{5}$$

To understand the form of (5), start by supposing $\mathbf{S}_{t+h\mid t}$ is a diagonal matrix, with the diagonal consisting of the cross-sectional variances $\left\{S_{jj,t+h\mid t}\right\}$, for j=1, . . . , n, where n is the number of variables. Then the measure simplifies to:

$$D_{i,t+h\mid t} = \sqrt{\sum_{j=1}^{n} \frac{\left(\mathbf{y}_{j,i,t+h\mid t} - \overline{\mathbf{y}}_{j,t+h\mid t}\right)^2}{S_{jj,t+h\mid t}}} \tag{6}$$

That is, it is the sum of individual i's squared deviations of each variable from the consensus, divided by the cross-sectional variance. When n=1, the measure further simplifies to the distance between the forecast and consensus, divided by the cross-sectional standard deviation. But when n>1, and $\mathbf{S}_{t+h|t}$ is non-diagonal, the value of the multivariate measure will be reduced (increased) when the vector of differences of individual i's forecasts is consistent with (at odds with) the interdependencies as measured by the cross-sectional forecast covariance matrix.

The univariate and multivariate individual measures of disagreement in (3) and (6) permit analysis of the heterogeneity and persistence of individual disagreement that are masked at the aggregate level.

c. Density Forecasts

Background and General Discussion

Many surveys of professional forecasters do not provide information about the degree of confidence that respondents attach to their point forecasts. There are, however, a smaller number of surveys, including the US-SPF and ECB-SPF, that report both point and density forecasts. Density forecasts provide a basis for constructing measures of uncertainty and studying their properties. While density forecasts are not the only source for measures of uncertainty, they have potential advantages compared to model-based measures, such as those derived from time series conditional variance models (following the seminal contribution by Engle (1982)). As noted by Giordani and Söderlind (2003), model-based estimates of uncertainty may lag those derived from surveys when there are structural breaks.

The implementation of density forecasts, however, can also introduce complications. For example, the ECB-SPF 2009:Q1 one-year-ahead real GDP growth forecasts witnessed a "pile-up" of probabilities in the lower open interval because the survey design failed to provide sufficient histogram coverage to accommodate respondents' shift toward a very pessimistic outlook for growth. Consequently, it can be especially challenging to calculate uncertainty estimates for GDP growth from the density forecasts for this specific survey.

Another complication is that the number, location, and widths of the bins used to define the histogram for the density forecasts occasionally change over time. In the case of the GDP-deflator inflation density forecasts for the US-SPF, the survey initially offered 15 interior bins with a width of 1% (1968:4 to 1981:2), with the locations of the bins changing in 1973:2 and 1974:4. From 1981:3 to 1991:4, the survey presented respondents with only 6 interior bins with a width of 2% (and a change in locations in 1985:2). From 1992:1 to 2013:4, the histogram consisted of 10 interior bins with a 1% width, and from 2014:1 the width was reduced to half a percentage point.[9] This type of change to the structure of the survey instrument makes it problematic for some measures derived from the density forecasts to have a consistent interpretation over time.

Uncertainty and Density Forecasts

Uncertainty reflects the confidence attached to a prediction and is determined from the distribution of probabilities that a forecaster attaches to the different possible outcomes (or the outcomes to different percentiles) of the target variable. In the case of the surveys, this is typically reported in the form of a histogram where increases (decreases) in the spread of the reported predictive probabilities are indicative of higher (lower) uncertainty.

Uncertainty at the Individual Level

Surveys that elicit density forecasts typically ask respondents to assign a sequence of probabilities, $p_i(k)$, to a set of $k = 1, \ldots, K$ pre-specified outcome intervals where each closed

[9] Recently, the COVID-19 pandemic resulted in marked changes in some of the histograms for the US-SPF and ECB-SPF starting with the 2020:Q2 survey. For example, lower and upper open limits were pushed out, and the histogram widths were made much wider in the tails than around the (assumed) center of the distribution. Because survey operators may not be able to anticipate every occasion when respondents will significantly shift their density forecasts, an alternative survey design would be to ask respondents to provide selective percentiles of their subjective distributions, but this is a complex issue (see, e.g., O'Hagan et al. (2006)).

interval is characterized by a lower bound, l_k, and an upper bound, u_k.[10] To facilitate discussion, we assume that the number of bins and the number of respondents do not change over time, and therefore, we omit a time subscript for these terms here as well as going forward.

One approach to calculating the mean and uncertainty associated with a density forecast is to use formulas for the mean and variance of a histogram. This approach, however, requires addressing two issues. The first is that density forecasts typically contain open intervals on each end of the histogram that, if the respondent assigned a probability to either, must be closed to calculate the mean and variance. A typical—although ad hoc—practice is to assign the same or twice the width of the interior closed intervals to the open intervals. The degree to which this will impact any estimate depends on the amount of assigned probability to each open interval.

The second issue concerns the location of the probability mass within a specified closed interval. For example, some studies [Boero, Smith, and Wallis (2008a), Rich and Tracy (2010), Kenny, Kostka, and Masera (2015), and Poncela and Senra (2017)] assume that the probability mass is located at the midpoint of each bin. Under this assumption, the mean and variance of forecaster *i*'s histogram are given, respectively, by:

$$\mu_{i,t+h\,|\,t} = \sum_{k=1}^{K} p_{i,t+h\,|\,t}(k) \times mid(k) \tag{7}$$

$$\sigma^2_{i,t+h\,|\,t} = \sum_{k=1}^{K} p_{i,t+h\,|\,t}(k) \times [mid(k) - \mu_{i,t+h\,|\,t}]^2 \tag{8}$$

where $p_{i,t+h\,|\,t}(k)$ is the probability that forecaster *i* assigns to the *k*th interval corresponding to an *h*-step-ahead prediction from a survey conducted in period *t* and $mid(k) = (u_k + l_k)/2$ is the midpoint of the *k*th bin.

Several studies [Zarnowitz and Lambros (1987), Abel et al. (2016), and Rich and Tracy (2021a)] instead assume that the probability is distributed uniformly within each interval. While the mean of the histogram derived under this assumption is identical to that in (7), the variance is now given by:

[10] The probabilities are assumed to sum to unity.

$$\sigma^2_{i,t+h\,|\,t} = \left[\sum_{k=1}^{K} p_{i,t+h\,|\,t}(k) \times \left(\frac{u_k^3 - l_k^3}{3(u_k - l_k)}\right)\right] - \left[\sum_{k=1}^{K} p_{i,t+h\,|\,t}(k) \times \left(\frac{u_k^2 - l_k^2}{2(u_k - l_k)}\right)\right]^2 \tag{9}$$

Recent work by Glas (2020) suggests that the variance estimates are not particularly sensitive to using method (8) or (9).

An alternative approach is to first fit a continuous parametric distribution to the individual histogram. This practice involves selecting a respondent-specific parameter vector $\theta_{i,t,h}$ for a proposed distribution that solves the following minimization problem:

$$\min_{\theta_{i,t,h}} \sum_{k=1}^{K} \left[F(u_k, \theta_{i,t,h}) - P_{i,t+h\,|\,t}(k) \right]^2 \tag{10}$$

where F is the cumulative distribution function (CDF) of the proposed distribution and $P_{i,t+h\,|\,t}(k) = \sum_{s=1}^{k} p_{i,t+h\,|\,t}(s)$ is the CDF corresponding to the reported histogram. Depending on the proposed distribution, the mean and variance of the histogram are obtained either directly or indirectly via the estimated parameter vector $\hat{\theta}_{i,t,h}$.

Most studies adopt either the normal distribution [Giordani and Söderlind (2003), Lahiri and Liu (2006), D'Amico and Orphanides (2008), Söderlind (2011), and Boero, Smith, and Wallis (2015)] or the generalized beta distribution with fixed support [Engelberg, Manski, and Williams (2009), Glas and Hartman (2016), Liu and Sheng (2019), and Glas (2020)]. Other choices include a gamma distribution [D'Amico and Orphanides (2008)] and a two-piece normal distribution [Krüger (2017)]. Importantly, Engelberg, Manski, and Williams (2009) note that the approach in (10) requires forecasters to assign non-zero probabilities to at least three bins. If a respondent assigns probability only to one or two bins, they suggest fitting triangular distributions and outline a method of doing so for two-bin histograms.

The IQR provides an alternative measure of individual uncertainty.[11] Following Abel et al. (2016) and Glas (2020), we can define this measure as:

[11] The IQR is typically more robust than a standard deviation/variance estimate to situations when respondents place probability in open intervals. The IQR is unaffected unless the respondent places more than a 25% probability in an open interval.

$$p_{i,t+h\mid t}^{IQR} = p_{i,t+h\mid t}^{0.75} - p_{i,t+h\mid t}^{0.25} \tag{11}$$

where $p_{i,t+h\mid t}^{0.75}$ and $p_{i,t+h\mid t}^{0.25}$ denote the 75th and 25th percentiles, respectively, of respondent *i*'s density forecast. The quartiles in (11) can be derived under the "mass-at-midpoint" and uniform distribution assumptions, as well as the parametric approach.[12]

Uncertainty at the Aggregate Level

A measure of aggregate uncertainty can be calculated as the mean of the standard deviations calculated from the individual density forecasts (see, e.g., Zarnowitz and Lambros (1987)):

$$\bar{\sigma}_{t+h\mid t} = N^{-1}\sum_{i=1}^{N}\sqrt{\sigma_{i,t+h\mid t}^{2}} \tag{12}$$

Batchelor and Dua (1996) and Boero, Smith, and Wallis (2008b, 2015) adopt the root mean subjective variance (RMSV) of the individual variances:

$$RMSV_{t+h\mid t} = \sqrt{(1/N)\sum_{i=1}^{N}\sigma_{i,t+h\mid t}^{2}} \tag{13}$$

An IQR-based measure of aggregate uncertainty has also been proposed based on an ordered array of the individual IQR values in (11). Abel et al. (2016) and Glas (2020) consider the cross-sectional median value for the uniform distribution and parametric approach. In the case of the "mass-at-midpoint" approach, Glas (2020) argues for the use of the cross-sectional average.

The individual density forecast data can be combined into an aggregate density forecast, from which various summary statistics such as the mean and standard deviation/variance can be calculated.[13] Wallis (2005) demonstrates that a measure of aggregate uncertainty can be obtained as the difference between the variance of the aggregate density forecast and dispersion in the individual point forecasts. Using the finite mixture distribution to characterize the statistical properties of the

[12] Under the "mass-at-midpoint" assumption, the quartiles in (11) are defined as the midpoint of the bin where the empirical CDF, $P_{i,t+h\mid t}(k)$, first exceeds the relevant threshold. In the case of the uniform distribution, linear interpolation is used to derive the quartiles.

[13] The aggregate density forecast typically is constructed as an equally weighted combination of the individual density forecasts.

aggregate density forecast, Wallis (2005) derives the following decomposition:

$$\begin{aligned} VAR\left(\phi_{t+h \mid t}\right) &= N^{-1}\sum_{i=1}^{N_t}\sigma^2_{i,t+h \mid t} + N^{-1}\sum_{i=1}^{N_t}\left(y_{i,t+h \mid t} - \overline{y}_{t+h \mid t}\right)^2 \\ &= \overline{\sigma}^2_{t+h \mid t} + \tilde{s}^2_{t+h \mid t} \end{aligned} \tag{14}$$

where $VAR\left(\phi_{t+h \mid t}\right)$ is the variance of the aggregate density forecast, $\overline{\sigma}^2_{t+h \mid t}$ is the average variance of the individual density forecasts, and $\tilde{s}^2_{t+h \mid t}$ is a measure of the cross-sectional variance of the individual point forecasts. As shown in (14), the variance of the aggregate density forecast reflects a combination of both average uncertainty and disagreement. Consequently, the spread of the aggregate density forecast should not be interpreted as a measure of aggregate uncertainty due to the additional influence of forecast dispersion.[14]

Given estimates of $VAR\left(\phi_{t+h \mid t}\right)$ and values of $\tilde{s}^2_{t+h \mid t}$, measures of aggregate uncertainty $\overline{\sigma}^2_{t+h \mid t}$ can be backed out from (14). An advantage of the approach is that the fitting of parametric distributions is facilitated by the fact that non-zero probabilities are almost always assigned to more than three bins of the aggregate density forecast. Boero, Smith, and Wallis (2008b) and Rich and Tracy (2010) adopt this approach for their analyses of the Bank of England Survey of External Forecasters (BOE-SEF) and the US-SPF, respectively.

Skew/Risk

Density forecasts also convey the shape of the subjective probability distribution and thereby provide an assessment about the respondent's balance of risks. That is, a respondent can report a symmetric histogram suggesting balanced risks to a forecast or a skewed histogram indicating upside or downside risks. Taken together, these additional features of density forecasts are important to policymakers and market participants as they allow for a richer characterization of forecasts and offer a more informed outlook.

d. A Closer Look at Disagreement and Uncertainty

Heterogeneity and Persistence at the Individual Level

---

[14] Another problematic aspect associated with using the spread of the aggregate density forecast as a measure of aggregate uncertainty is that Wallis (2005) shows that changes in the variance of aggregate density forecasts principally reflect shifts in disagreement and not individual uncertainty.

Aggregate measures of expectations and/or uncertainty from surveys cannot speak to issues related to the forecast behavior of individual respondents in the cross-section or over time. Consequently, panel data on individual forecasts are attractive for the construction of key variables of interest and for assessing differences across respondents and the statistical properties of those differences.

Underlying the earlier discussion of measures of disagreement is the observation that there is marked dispersion across the cross-sectional distribution of point forecasts. Turning to the individual density forecasts, various studies [D'Amico and Orphanides (2008), Boero, Smith, and Wallis (2008a, 2008b, 2015), Clements (2014b), Glas (2020), Rich and Tracy (2021a)] also find considerable heterogeneity in the cross-sectional distributions of the first- and second moments of the density forecasts. That is, respondents display notable differences in the mean and confidence of their reported outlooks.

The notion of forecaster disagreement can be extended to their density forecasts. Specifically, Rich and Tracy (2021a) propose an individual density disagreement measure using the Wasserstein distance (WD) measure for histograms (see Arroyo and Mate (2009)). Specifically, the WD disagreement measure between respondent *j* and respondent *i* is given by:

$$WD_{j,i,t+h\mid t} = \int_0^1 \left| F^{-1}_{j,t+h\mid t}(z) - F^{-1}_{i,t+h\mid t}(z) \right| dz \tag{15}$$

where $F^{-1}_{i,t+h\mid t}$ denotes the inverse CDF for respondent *i* in the survey at date *t*.[15] They then define the average absolute density disagreement (AADD) measure for respondent *i* as follows:

$$AADD_{i,t+h\mid t} = \frac{1}{N-1}\sum_{j\neq i} WD_{j,i,t+h\mid t} \tag{16}$$

The expression in (16) has several attractive features. First, the measure uses the same density forecasts as those used for the individual uncertainty measure. Second, the measure captures any differences across respondents' density forecasts by focusing on the entire cumulative distribution function.

[15] The squared distance norm can be used as an alternative to the absolute value norm in (15).

As previously discussed, the evidence from cross-sectional distributions indicates heterogeneity in respondents' point and density forecasts. However, the heterogeneity does not reveal the degree to which respondents move within the distributions over time. In particular, the distributions do not indicate whether there are persistent patterns in individual respondents' forecast behavior and, if so, the sources for such persistence. The availability of measures of uncertainty and disagreement at the individual level allows for an exploration into the issue of persistence for these features of forecast behavior.

On the confidence of respondents' predictions, Boero, Smith, and Wallis (2015) and Rich and Tracy (2021a) find strong evidence of persistence in individual uncertainty. Moreover, Rich and Tracy (2021a) find that individual uncertainty appears to be associated with a prominent respondent effect. That is, while there are marked differences across forecasters in the confidence attached to their predictions, forecasters' confidence changes slowly over time.

Turning to disagreement, Clements (2021b) adopts the individual multivariate disagreement measure in (5) to determine whether some individuals in the US-SPF are persistently more contrarian than others. Using a split sample analysis, he ranks forecasters in each period based on the extent of their disagreement. His results document a positive correlation between the ranks in the two periods, indicating persistence in disagreement. Rich and Tracy (2021a) examine data from the ECB-SPF and adopt the personal disagreement measure in (3). For each survey period, they assign forecasters to quartiles based on the extent of their disagreement and then examine the transitions between quartiles over time. The results indicate that forecasters tend to remain in the same quartile, also indicating persistence in disagreement. In contrast to their analysis of uncertainty, Rich and Tracy (2021a) find that individual disagreement is associated with a prominent time effect.

Taken together, individual data reveal notable heterogeneity and persistence in uncertainty and disagreement. Moreover, this evidence is consistent with other work documenting heterogeneity and persistence in the relative level of point forecasts [Patton and Timmermann (2010), and Boero, Smith, and Wallis (2015)]. These properties do not support full information rational expectations (FIRE) models, which cannot generate heterogeneity, nor models that incorporate informational rigidities, such as Coibion and Gorodnichenko [(2012), (2015b)]. At least in their baseline versions, these latter models assume homogeneous signal qualities (noisy information) or propensities to update (sticky information) and therefore cannot generate systematic differences across forecasters.

Rather, the observed properties of individual forecast behavior cited above would argue for the development of expectations models that feature persistent heterogeneity.

Is Disagreement a Reliable Proxy for Uncertainty?

Empirical studies attempting to quantify the effects of uncertainty face the challenge of constructing measures of uncertainty.[16] While researchers may find appropriate density forecasts from surveys, absent this situation they will need to select a proxy to measure uncertainty. In practice, the dispersion of point forecasts has been used for this purpose. However, the validity of this practice requires a significant positive association between disagreement and uncertainty. Zarnowitz and Lambros (1987) show that it is possible for disagreement and uncertainty to display a negative relationship or no relationship. Essentially, for this reason, Manski (2011) and others are critical of the use of disagreement to proxy uncertainty.

Whether there is a meaningful correlation between disagreement and uncertainty is an empirical question and several studies have investigated this relationship. These studies, however, have largely been conducted at the aggregate level and have relied mainly on data from the US-SPF. The following regression model is typically used to assess the association between uncertainty and disagreement:

$$Uncertainty_{t+h\,|\,t} = \alpha + \beta(Disagreement_{t+h\,|\,t}) + \varepsilon_{t+h\,|\,t} \tag{17}$$

where the measures of aggregate uncertainty and disagreement (expressed in comparable units) are selected from the various candidate series previously discussed. The issue of whether disagreement is a valid proxy for uncertainty rests on both statistical and economic significance. Specifically, the determination is based on a one-tailed test that $\beta > 0$ and the R^2 from the estimation of (17).

The evidence from the US-SPF has been mixed. Zarnowitz and Lambros (1987) report a modest positive association between disagreement and uncertainty. Giordani and Söderlind (2003)

[16] There is an extensive literature investigating the effects of uncertainty on economic activity and financial variables. Early work focused on the dynamics of uncertainty and capital investment [Bernanke (1983), Dixit (1989), Abel and Eberly (1994), and Dixit and Pindyck (1994)], with more recent work by Leahy and Whited (1996) and Bloom, Bond, and van Reenen (2007) linking lower investment to stock market volatility. In addition, uncertainty can affect the hiring decisions of firms [Bloom (2009), Schaal (2017)] and serves as an important source for business cycle fluctuations [Bloom et al. (2018)]. Campbell (2000) discusses how uncertainty is a fundamental determinant of asset prices. Turning to the international sphere, uncertainty can affect export dynamics [Novy and Taylor (2020)] or the decisions of firms to enter and invest in new export markets [Handley and Limão (2015, 2017)].

extend the sample period of Zarnowitz and Lambros and report a positive association between disagreement and uncertainty that is both economically and statistically significant, although some studies have argued that their conclusion is problematic.[17] Examining matched point and density forecasts from the US-SPF Rich and Tracy (2010) find a very weak relationship between disagreement and uncertainty. Recent analyses that have examined data from other surveys featuring point and density forecasts, such as Boero, Smith, and Wallis (2008b) for the BOE-SEF and Abel et al. (2016) and Glas (2020) for the ECB-SPF, have found little support for the use of disagreement as a proxy for uncertainty.[18]

Lahiri and Sheng (2010) suggest the disagreement-uncertainty relationship may be episodic and that the strength of the disagreement-uncertainty relationship is inversely related to the volatility of the forecasting environment. Lahiri and Sheng find some support using the US-SPF data. Rich and Tracy (2021a) construct measures of disagreement and uncertainty at the individual level for the ECB-SPF and use a panel data model to test the relationship between the measures and any changes in the strength of the relationship across time. They find an economically insignificant relationship between the variables in both the pre- and post-financial crisis periods, with their results corroborating the preponderance of evidence at the aggregate level that disagreement is not a reliable proxy for uncertainty. Moreover, they argue that the weak linkage between uncertainty and disagreement could reflect the divergent properties of the variables discussed in the preceding section in which uncertainty is characterized by prominent respondent effects and disagreement is characterized by prominent time effects.

Ex Ante vs. Ex Post Uncertainty

A histogram-based measure of uncertainty is an *ex ante* measure in that it is available before the realization of the target variable becomes known. *Ex post* measures of forecast uncertainty, such as the root mean squared forecast error (RMSFE), only become available when the realizations are known. A number of studies have sought to compare the two uncertainty measures as a way to address questions relating to whether respondents' *perceptions* of uncertainty are accurate, that is,

17 Rich and Tracy (2010) and Boero, Smith, and Wallis (2015) discuss the problems with Giordani and Söderlind (2003) fitting normal distributions to two-bin histograms. Rich and Tracy (2010) also raise concerns about deriving a disagreement measure from the (estimated) means of the density forecasts rather than the point forecasts.

18 Boero, Smith, and Wallis (2015) find a strong positive correlation between disagreement and uncertainty when they extend their 2008 analysis of the BOE-SEF to include the global financial crisis period. However, they note that the results are largely driven by observations associated with the initial onset of the crisis.

whether the *ex ante* and *ex post* measures broadly match.

Clements (2014b) looks at both individual and aggregate forecasts for the US-SPF, exploiting the point and histogram forecasts to calculate perceived (*ex ante*) and realized, or actual, (*ex post*) uncertainty. He finds that respondents tend to be under-confident in their probability assessments of both inflation and GDP growth at horizons up to one year ahead. For forecasts of the *next* year the probability assessments display over-confidence.

A similar dependence on the forecast horizon (or the "term structure") is found by Knüppel and Schultefrankenfeld (2019). In their study of central banks' inflation uncertainty forecasts (they consider the Bank of England, the Banco Central do Brasil, the Magyar Nemzeti Bank, and the Sveriges Riksbank), they find that the banks' forecasts also tend to display under-confidence at short horizons and over-confidence at longer horizons.

III. Evaluation of Forecaster Performance

a. Data Revisions

Most of the target variables in surveys of professional forecasters are subject to revision over time. In these cases, which vintage of data should be taken as the "actual values" for calculating forecast errors, given that the analyst is ignorant about the respondent's choice of vintage? A forecast might be reasonably close to the advance estimate but end up deviating considerably from a revised estimate. This problem is obviously dependent on the size of the revisions. For example, Croushore (2011a, p.249, Figure 9.1) illustrates, with an admittedly extreme case, the revisions to the 1976:Q3 growth rate of U.S. real residential investment. A plot of the initial release and of all subsequent estimates (up to those made in 2009) show that the estimates of the annual rate rose from less than 3% to nearly 16%, before ending up at around −5%!

Using later estimates for actual values might be advised on the grounds that these are more accurate estimates of the "true" values, which are of primary interest, and *ought* to be the focus of respondents. On the other hand, Keane and Runkle (1990, p. 715) argue that the use of revised data in tests of forecaster rationality might lead to misleading results, since it rests on incorrect assumptions "about what the forecasters tried to predict and what they knew when they made their predictions."

In addition to the magnitude of revisions, a key consideration is whether revisions are predictable (see, e.g., Mankiw and Shapiro (1986)). Clements and Galvão (2017b) show that revisions can sometimes be predicted using general information sets, and Clements (2019) shows that under certain conditions it is possible in principle to determine the vintage of the data being targeted.

b. Rationality/Efficiency of Point Forecasts

Starting with Turnovsky and Wachter (1972), both aggregate and individual survey expectations have been used to determine whether economic agents form their expectations rationally.

Rational expectations is the hypothesis that:

> "expectations, since they are informed predictions of future events, are essentially the same as the predictions of the relevant economic theory" (Muth (1961, p. 316)).

This statement assumes that all agents have access to all relevant information. Several authors have balked at this assumption (see, e.g., Pesaran (1987)), and it may be more reasonable to ask whether agents make efficient use of the information they possess. This more limited notion of forecast efficiency proposed by Mincer and Zarnowitz (1969) can be tested by regressing actual values on forecasts:

$$y_t = \alpha_i + \beta_i y_{i,t\,|\,t-h} + \mu_{i,t} \tag{18}$$

where $y_{i,t\,|\,t-h}$ is the forecast made by individual *i* at $t-h$ of the target variable y_t, and the data range over *t* for a fixed horizon *h*. The null of forecast efficiency is that $\alpha_i = 0$ and $\beta_i = 1$. Consider the covariance between the forecast error and the forecast:

$$Cov\left(y_t - y_{i,t\,|\,t-h}, y_{i,t\,|\,t-h}\right) = Cov\left(\left(\beta_i - 1\right) y_{i,t\,|\,t-h} + \mu_{i,t}, y_{i,t\,|\,t-h}\right) \tag{19}$$

Unless $\beta_i = 1$, the covariance will be non-zero and could be exploited to adjust the forecasts to improve their accuracy. Hence, the original forecasts do not make optimal use of the information available to the forecaster, where "information" in this context simply refers to the respondent's own forecast. (If $\alpha_i \neq 0$ when $\beta_i = 1$, the forecasts are clearly biased.) For multi-step forecasts, $h > 1$, heteroscedasticity, and autocorrelation consistent (HAC) standard errors are used to account for the overlapping forecasts phenomenon, which induces serial correlation in $\mu_{i,t}$ in (18).

Should the test for forecast efficiency be applied to aggregate or individual forecasts? The aggregate may appear a better choice, since it avoids the vagaries of any given forecaster, but the comment on Figlewski and Wachtel (1981) by Dietrich and Joines (1983), and the rejoinder by

Figlewski and Wachtel (1983), establish that for the aggregate, $\beta < 1$ (when $\beta_i = 1$ for all *i,* unless the forecasts are identical across individuals). That is, an aggregate of forecast-efficient individuals will appear to be inefficient.

A related finding is the under-reaction of the consensus forecasts to new information, as predicted by models of dispersed noisy information or rational inattention. This literature regresses the aggregate forecast error on the revision to the aggregate forecast and shows that the slope coefficient should be positive. Individual forecasters correctly down-weight their information because it is noisy, but this results in an under-response of the average forecasts. Regressions of this type have recently generated much interest (see Coibion and Gorodnichenko (2015b), Broer and Kohlhas (2018), Bordalo et al. (2020) and Angeletos, Huo, and Sastry (2020), inter alia).

More demanding tests of rationality would require the forecast error to be orthogonal to variables in the agent's information set at time $t-h$. A term such as $g'_{i,t-h}\theta_i$ could be included in (18), where $g_{i,t-h}$ denotes a vector of variables in the agent's information set at time $t-h$. The null specified above is then augmented by the requirement that $\theta_i = 0$.

As noted in Section II, some surveys have a fixed-event dimension. Patton and Timmermann (2012) propose a test that exploits this dimension of the forecast data by essentially testing whether the long-horizon forecast error is systematically related to revisions to more recent forecasts of the target variable (their "optimal revision regression"). The concern that inference may be affected by an erroneous choice of the outcome vintage can be countered by replacing the realization of the target variable by a short-horizon forecast. In this *Handbook,* Chapter 17: "The Term Structure of Expectations" presents a model of the relationship between different horizon forecasts.[19]

Tests of efficiency or rationality might fail to detect alternating periods of under-prediction and over-prediction, such as the tendency of U.S. inflation forecasts to under-predict in the 1970s but to over-predict in the 1980s and 1990s. Rossi and Sekhposyan (2016) use the US-SPF inflation forecasts and the Federal Reserve's Greenbook forecasts to test for forecast optimality allowing for instabilities. The null hypothesis of rationality is not rejected in the standard setting, but rationality is rejected for both sets of forecasts after allowing for instabilities.

c. Scoring Rules

[19] These tests are closely related to the weak efficiency tests of Nordhaus (1987) and the property that forecast revisions should be unpredictable from earlier revisions.

Point Forecasts

In practice, point forecasts are often evaluated using the absolute or squared forecast error averaged over the set of forecasts and realizations. In Section IV, we discuss loss functions for point forecasts that do not penalize equal under- and over-predictions the same. Allowing for asymmetric loss might be viewed as a small step in the direction of evaluating forecast errors in terms of economic costs (see, e.g., Granger and Machina (2006)).

Density Forecasts

A large and growing body of work focuses on the properties of density forecasts: their accuracy, whether they are calibrated (in the sense of closely matching the true densities), and whether survey participants' reported density forecasts are overly optimistic or pessimistic.

Density forecast accuracy is typically evaluated using a scoring rule, which can also be used to rank competing forecasts. In line with the histogram format of the survey instruments considered here, we can describe this approach for respondent *i* at time *t* for a target variable at time *t* + *h* as follows. We assume that there are K bins associated with the histogram at time *t* and let $p_{i,t+h\,|\,t}(k)$ denote the probability assigned by respondent *i* to the *k*th bin. While there are several candidates for a density-based accuracy measure, the ranked probability score (RPS) [Epstein (1969)] is a popular choice and can be defined as follows:

$$RPS_{i,t+h\,|\,t} = \frac{1}{K-1}\sum_{j=1}^{K}\left(\sum_{k=1}^{j} p_{i,t+h\,|\,t}(k) - \sum_{k=1}^{j} I_{t+h}(k)\right)^2 \tag{20}$$

where $I_{t+h}(k)$ denotes an indicator variable that takes a value of one if the actual outcome in period *t*+*h* is in the k^{th} interval of the histogram from the survey at date *t*. The forecast performance metric in (20) can then be averaged over all surveys in which respondent *i* participated to produce an overall score. Boero, Smith, and Wallis (2011) provide an overview of scoring rules.

d. Using the Probability Integral Transform to Assess Density Forecast Coverage

Density forecasts can also be evaluated using the probability integral transform (PIT), as surveyed in Diebold, Gunther, and Tay (1998) (see also Dawid (1984) and Rosenblatt (1952)). Let $\{f_t(y_t)\}_{t=1}^{m}$ be the sequence of true conditional densities governing the target variable y_t, $\{y_t\}_{t=1}^{m}$

the sequence of realizations of y_t, and let $\left\{\Omega_t\left(y_t\right)\right\}_{t=1}^{m}$ be the corresponding sequence of one-step-ahead density forecasts. Then z_t is the series of probability integral transforms defined by:

$$\left\{z_t\right\}_{t=1}^{m}=\left\{\int_{-\infty}^{y_t}\Omega_t\left(u\right)du\right\}_{t=1}^{m} \tag{21}$$

A sequence of density forecasts is consistent with the data-generating process for the target variable, i.e., $\left\{\Omega_t\left(y_t\right)\right\}_{t=1}^{m}=\left\{f_t\left(y_t\right)\right\}_{t=1}^{m}$, if $\left\{z_t\right\}_{t=1}^{m}$ is i.i.d. $U(0,1)$. Various testing procedures are available. Berkowitz (2001) suggests taking the inverse normal transformation of $\left\{z_t\right\}_{t=1}^{m}$, and testing whether this is an i.i.d drawing from a standard normal, using likelihood ratio tests based on Gaussian likelihoods. See also Knüppel (2015) for an alternative approach.

Two sets of forecast densities can also be compared against each other. Bao, Lee, and Saltoglu (2007) show that the difference between the Kullback-Leibler information criterion (KLIC) calculated for two rival forecast densities can be used as the basis for a test of equal predictive ability using the Diebold and Mariano (1995) approach. (A similar approach has been developed by Amisano and Giacomini (2007).)

The evidence concerning the coverage rates of respondents' density forecasts is mixed. For example, Diebold, Tay, and Wallis (1999) and Giordani and Söderlind (2003, 2006) examine the US-SPF and find that forecasters are "over-confident" when reporting their probabilistic assessment of target variables. Kenny, Kostka, and Masera (2014) and Krüger (2017) find similar results examining the ECB-SPF, while Clements (2004) examines the Bank of England Inflation Reports and finds the one-year-ahead forecasts unduly pessimistic in that too much probability mass is allocated to high rates of inflation, whereas the current and next quarter forecasts are better calibrated.

Clements (2018) compares the US-SPF GDP deflator inflation and real GDP growth histograms against a set of benchmark forecasts, both at an individual level and in the aggregate, using the approach of Bao, Lee, and Saltoglu (2007). The benchmark forecasts are constructed from the SPF historical median forecast errors and are unconditional in nature. The SPF forecasts are generally not more informative than the benchmark forecasts.

e. Balanced vs. Unbalanced Panels

Missing observations from non-response complicate individual-level analyses of forecaster behavior, as well as the calculation of aggregate measures. Due to the unbalanced panel structure, analyses of individual-level forecast performance may need to control for distortions that arise from respondents participating at different times, especially if these times are characterized by different economic conditions.

Calculating an aggregate measure from a series of individual responses can be regarded as a forecast combination problem. That is, the optimal combination will depend on the loss function, and the weights attached to the individual forecasts will depend on the relative accuracy of the individual forecasts and the correlations between the individual forecast errors.[20]

In practice, the cross-sectional median or mean ("equal weights") is routinely used. Genre et al. (2013) consider whether it is possible to do better than "the simple average" for the ECB-SPF. As noted by Capistrán and Timmermann (2009), missing forecast values imply there may be relatively few observations on which to estimate optimal combination weights. Uncertainty in the optimal weights might favor the use of theoretically sub-optimal equal weighting of the individual forecasts. Capistrán and Timmermann (2009) consider "filling in" missing data before combining. They use the EM algorithm and a simple model for individual forecasts to back-fill missing forecasts, although such approaches are not common in the survey expectations literature.

Engelberg, Manski, and Williams (2011) focus on how to interpret changes in the consensus forecast when the composition of the panel is changing over time. For example, compositional effects may mean that the aggregate changes in ways that are at odds with changes in the consensus of the ever-present forecasters. Engelberg, Manski, and Williams (2011, p. 1061) argue that changing panel composition can only be ignored "if it were credible to assume that panel members are randomly recruited from a stable population of potential forecasters and that participation in the survey after recruitment is statistically independent of forecasters' beliefs about inflation." They propose replacing missing forecasts with imputed upper and lower values, such that the implied changes between periods are no more extreme than the observed changes in the forecasts of those who responded in both periods. This generates a bound on the change in the aggregate. Consistent with the views

[20] See Bates and Granger (1969), Granger and Ramanathan (1984), and Clements and Harvey (2009) for a review. Although in general the optimal weights will differ for asymmetric loss, Elliott and Timmermann (2004) show that under certain conditions the weights are identical to the squared-error loss weights, and the degree of asymmetry is accommodated by the constant term in the combination.

expressed in Manski (2007), in this instance a more "credible" assumption concerning the missing data leads to a bound as opposed to a point estimate of the change in the aggregate.

Clements (2021a) considers whether joiners and leavers differ from incumbents for the US-SPF and finds some evidence that joiners' density forecasts are less accurate. However, it is difficult to gauge the reliability of this finding due to the relatively small number of joiners to each survey.

f. Are Some Forecasters Better Than Others?

Compared to households or firms, professional forecasters are often viewed as a homogeneous group. Nevertheless, researchers have investigated whether some forecasters are better/worse than others. Early studies by Stekler (1987) and Batchelor (1990) report contradictory findings for the Blue Chip survey, with the different conclusions related to the debate about the appropriate test statistic for the analysis. Christensen et al. (2008) examine the US-SPF and develop a test for equal forecasting accuracy based on the forecast comparison test of Diebold and Mariano (1995). Their analysis yields mixed results, with tests suggesting equal predictive performance for some variables and not others. However, they are only able to study three individual respondents because their approach requires a balanced panel and a long time series of forecasts.

D'Agostino, McQuinn, and Whelan (2012) point out that there are notable drawbacks to the approaches used in these previous studies. First, the requirement of a balanced panel can significantly reduce the sample size. Second, when respondents participate at different times, inter-respondent comparisons need to control for differences over time in the forecasting environment. To do this, they suggest using a normalized forecast error. If $e_{i,t+h\mid t}$ is the forecast error made by individual i forecasting y_{t+h} at survey date t, the normalized forecast error is:

$$\tilde{e}_{i,t+h\mid t} = \frac{e_{i,t+h\mid t}}{\sqrt{\frac{1}{N_t}\sum_{j=1}^{N_t} e^2_{j,t+h\mid t}}} = \frac{e_{i,t+h\mid t}}{\sqrt{\left(\overline{e^2_{t+h\mid t}}\right)}} \tag{21}$$

where N_t is the number of respondents at survey date t.

D'Agostino, McQuinn, and Whelan (2012) compare the empirical distribution of forecaster performance (using the normalized errors) to a simulated distribution calculated under the null hypothesis of equal ability, constructed by randomly assigning forecasts to forecasters. Their approach also accounts for the unbalanced nature of the panel. They find little evidence to suggest that some US-SPF forecasters are notably better than others, once a set of poorly performing forecasters is removed. Meyler (2020) finds similar results for the ECB-SPF. However, Clements (2020) finds systematic differences in the accuracy of the US-SPF histogram forecasts.

Rich and Tracy (2021b) use a regression approach to examine forecasts from the ECB-SPF. Specifically, they adopt a panel data specification for each respondent *i* and survey in period *t:*

$$FP_{i,t+h|t} = \alpha_i + \lambda_i \left(\overline{FP_{t+h\,|\,t}} \right) + \varepsilon_{i,t+h\,|\,t} \tag{22}$$

where $FP_{i,t+h|t}$ and $\overline{FP_{t+h\,|\,t}}$ denote a forecast performance (FP) metric at the individual and average (cross-sectional) level, respectively, and $\varepsilon_{i,t+h|t}$ is a mean-zero error term. An important feature of (22) is that it allows for two sources of heterogeneity in forecast performance across survey respondents through a fixed effect (α_i) and an individual loading factor (λ_i). They find strong evidence that respondents do not display equal forecast accuracy. Moreover, their results indicate that some respondents display higher relative accuracy in more tranquil environments, while others display higher relative accuracy in more volatile environments.

g. Professionals versus Models (and other Sources of Forecasts)

The predictions of professional forecasters have been compared to alternative sources of forecasts, including surveys of consumers and firms, as well as forecasts from empirical models, financial markets, and prediction markets (see, e.g., Snowberg, Wolfers, and Zitzewitz (2013) and Grothe and Meyler (2018) on the last two). Some earlier studies showed survey forecasts outperforming model forecasts for nominal variables such as inflation (see, e.g., Ang, Bekaert, and Wei (2007), Faust and Wright (2009, 2013), Aiolfi, Capistrán, and Timmermann (2011), and Clements (2015)), with more equivocal findings for real variables such as output. However, even for inflation, the evidence is mixed, with a recent study by Berge (2018) finding that an ARIMA(1,1) produces more accurate inflation forecasts than professional forecasters over 1990 to 2015.

A possible reason why survey-based forecasts might outperform those from models is that they are likely to benefit from a superior reading of the state of the economy and ongoing trends at the time the forecasts are made. They may also be able to react more rapidly than model-based forecasts to structural change. Against this, as discussed earlier, recent theories of expectations formation suggest that survey forecasts may not efficiently incorporate new information which might be expected to adversely affect accuracy.

A growing literature suggests using survey expectations in conjunction with empirical models. For example, Chan, Clark, and Koop (2018) use long-run inflation survey forecasts in a model of inflation, and Coibion, Gorodnichenko, and Kamdar (2018) argue for a greater use of survey data in macroeconomic analyses, with a focus on the New Keynesian Phillips curve. That said, Coibion and Gorodnichenko (2015a) argue for the primacy of firms' expectations over those of professional forecasters in Phillips curve models of inflation, and of using consumers' expectations to proxy firms' expectations when the latter are unavailable.

Some studies have compared surveys and other sources of forecasts in terms of density forecasts. Alternative sources of density forecasts are rarer, with the exception of model-based forecasts. In addition to Clements (2018) discussed earlier, there are comparisons of variances derived from the histograms to model forecasts. These include Giordani and Söderlind (2003) and Clements and Galvão (2017a). Clements and Galvão (2017a) found that model-based uncertainty assessments were more accurate than those of the US-SPF.

IV. Consistency of Point and Density Forecasts

When both point and density forecasts are available, a natural question is whether the two are consistent. We consider reasons why a respondent's forecasts may appear inconsistent and discuss formal evaluation for this property.

a. Calculating Bounds on the Central Moments of Histograms

Comparisons between point forecasts and density forecasts are complicated by the respondent only imperfectly revealing the underlying density forecast, and ignorance of what the point forecast represents (e.g., the mean of the subjective density, or the optimal forecast for an asymmetric loss function, and so on). Engelberg, Manski, and Williams (2009) suggest a way of dealing with the first consideration without introducing any auxiliary assumptions by calculating bounds on a given measure of the central tendency of the histogram (such as the mean, for example). If the point prediction lies within the bounds on a particular moment, such as the mean, then it is consistent with that moment of the underlying distribution.

The calculation of bounds on the mean, as well as the mode and median, can be explained by an example. Suppose a respondent assigns non-zero probabilities to three bins. The probability that inflation (say) is in the interval [2, 3] is 0.5, in the interval [3, 4] is 0.3, and is in the interval [4, 5] is 0.2. For the mean, the lower (upper) bound is obtained if all the probability is assumed to be at the lower (upper) limit of the histogram intervals. Then the lower bound is 0.5 x 2 + 0.3 x 3 + 0.2 x 4 = 2.7, and the upper bound is 0.5 x 3 + 0.3 x 4 + 0.2 x 5 = 3.7. Hence, a value of the point forecast in the interval [2.7, 3.7] could plausibly be interpreted as the respondent's (conditional) mean forecast. For the median, the bounds are given by the lower and upper values of the histogram intervals that contain 50% of the cumulative probability. But since the [2.0, 3.0] interval contains exactly 50% of the cumulative probability, the bound is increased to include the next bin; so for this histogram, the bound on the median is [2.0, 4.0]. The mode is given by the bin with the maximum probability; so here the bound on the mode is [2.0, 3.0].

Based on the percentages of point forecasts for the US-SPF (across respondents and surveys) that are within, below, and above the bounds calculated on the mean, median, and mode, Engelberg, Manski, and Williams (2009) conclude that point predictions may have a systematic, favorable bias in that when the point forecast does not lie within the bounds of any of the three central moments, the point forecasts are generally below the bounds for inflation and above the bounds for real GDP growth. However, an alternative interpretation is that the inaccuracy of the probabilistic forecasts is the source of this mismatch between the point predictions and the bounds (see Clements (2009, 2010, 2014a)).

b. Nature of Loss Functions – Symmetric vs. Asymmetric

Individuals may attach different costs to over- and under-predicting by the same amount. If their loss functions are asymmetric, then, as is well known, their forecasts will be biased (see, e.g., Granger (1969) and Zellner (1986)). The literature considers whether forecasts are rational once we allow forecasters to have asymmetric loss functions.

Under certain conditions on the data-generating process and the loss function, Patton and Timmermann (2007) show that the optimal forecast is given by:

$$y_{i,t+h\,|\,t} = \mu_{i,t+h\,|\,t} + \phi_{ih} \cdot \sigma_{i,t+h\,|\,t} \tag{23}$$

where $\mu_{i,t+h\,|\,t}$ is respondent *i*'s conditional mean, and $\sigma_{i,t+h\,|\,t}$ is the forecast standard deviation (or possibly the variance). The term ϕ_{ih} is a constant that depends on the form of the data-generating process and the loss function (e.g., the degree of asymmetry). As shown in (23), the conditional mean and optimal forecast will differ when point forecasts are optimal for an asymmetric loss function. This is a prima facie explanation of discrepancies between histogram means and point forecasts. The conditional mean and forecast standard deviation can of course be calculated from the histograms, provided we are prepared to make the sorts of assumptions that were avoided by the bounds approach.

Elliott, Komunjer, and Timmermann (2008) test for rationality allowing for a general loss function using the approach of Elliott, Komunjer, and Timmermann (2005). They find that rationality is rejected for fewer US-SPF respondents under asymmetric loss than when quadratic loss is imposed. However, Clements (2009, 2010) finds little evidence in favor of the asymmetric loss explanation for histogram/point forecast inconsistencies.

c. Rounding of Point and Density Forecasts

Although there is evidence that consumers round their (point) forecasts to a striking degree (see, e.g., Binder (2017)), there is much less evidence that professional forecasters round their point forecasts to the same extent (Clements (2021c)). The literature has focused on the rounding of probability forecasts by professionals based on the observed reporting in histograms and, in the case of the US-SPF, the probability of GDP decline forecasts. Engelberg, Manski, and Williams (2009, Appendix, pp. 40-1) consider whether rounding of the histogram probabilities could explain apparent inconsistencies between the histograms and point forecasts. They propose a strategy to "undo" any presumed rounding of the histograms, but find that their results concerning the patterns of inconsistency are qualitatively unchanged.

Clements (2011) considers whether allowing for "plausible" patterns of rounding behavior affects the finding that one in three pairs of probability forecasts of a decline in real output and histograms for annual real output growth is not consistent. One aspect of his approach is to replace a point probability of a decline by an interval, when there is evidence that an individual reports rounded forecasts. This is based on the assumption that a respondent applies the same rounding rule each time he or she responds to an SPF survey question on the probability of decline (c.f., Manski and Molinari (2010)). Allowing for rounding reduces the rate of inconsistent pairs of forecasts from one in three to one in four. While rounding has an effect, it is only a partial explanation of the observed inconsistencies.

V. Conclusion

Taken together, surveys of professional forecasters have provided critical insights into the expectations formation process. The surveys serve both to inform the development of expectations models and to test their implications, with the wide range of topics and extensive literature associated with the surveys indicating the extent of their contributions. The analysis of uncertainty afforded by the availability of density forecasts continues to draw considerable attention. Related issues such as the relationship between *ex ante* and *ex post* measures of uncertainty and the reliability of using forecast dispersion measures to proxy uncertainty remain areas of ongoing interest. The surveys also provide a basis for comparing the alignment of point forecasts and (estimated) means of the density forecasts as well as assessing the nature of forecasters' loss functions.

While surveys of professional forecasters have often been used for analyses at the aggregate level, interesting and important features emerge from examining the data at the individual level. For example, there is notable heterogeneity in forecast behavior that extends beyond first moments to include uncertainty, the degree of contrarianism, and accuracy. Moreover, and perhaps more importantly, there is growing evidence to suggest that this heterogeneity is persistent in nature. The possibility that these observed differences are systematic has important implications. At a minimum, this finding would call into question the view that professional forecasters can be treated as essentially identical and interchangeable, and future research should look to determine the source(s) of the persistent heterogeneity. This finding would also motivate studies to explore whether systematic differences in predictive behavior are unique to professional forecasters or if this characteristic extends to households and firms. If persistent heterogeneity is a general property, then this consideration would argue for the development of a new class of expectations models that are not only capable of generating differential forecast behavior but also providing mechanisms that can sustain such differences across time.

References

Abel, Andrew B., and Janice C. Eberly. 1994. "A Unified Model of Investment under Uncertainty." *The American Economic Review* 84 (5): 1369–84. https://ideas.repec.org/a/aea/aecrev/v84y1994i5p1369-84.html.

Abel, Joshua, Robert W. Rich, Joseph Song, and Joseph S. Tracy. 2016. "The Measurement and Behavior of Uncertainty: Evidence from the ECB Survey of Professional Forecasters." *Journal of Applied Econometrics* 31 (3): 533–50. https://doi.org/10.1002/jae.2430.

Aiolfi, Marco, Carlos Capistrán, and Allan Timmermann. 2011. "Forecast Combinations." In *The Oxford Handbook of Economic Forecasting*, 355–88. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195398649.013.0013.

Amador, Manuel, and Pierre-Olivier Weill. 2010. "Learning from Prices: Public Communication and Welfare." *Journal of Political Economy* 118 (5): 866–907. https://doi.org/10.1086/657923.

Amisano, Gianni, and Raffaella Giacomini. 2007. "Comparing Density Forecasts via Weighted Likelihood Ratio Tests." *Journal of Business & Economic Statistics* 25 (2): 177–90. https://doi.org/10.1198/073500106000000332.

Ang, Andrew, Geert Bekaert, and Min Wei. 2007. "Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?" *Journal of Monetary Economics* 54 (4): 1163–1212. https://doi.org/10.1016/j.jmoneco.2006.04.006.

Angeletos, George-Marios, Zhen Huo, and Karthik A. Sastry. 2021. "Imperfect Macroeconomic Expectations: Evidence and Theory." In *NBER Macroeconomics Annual 2020*, edited by Martin S. Eichenbaum and Erik Hurst, 35:1–86. https://doi.org/10.1086/712313.

Arroyo, Javier, and Carlos Maté. 2009. "Forecasting Histogram Time Series with K-Nearest Neighbours Methods." *International Journal of Forecasting* 25 (1): 192–207. https://doi.org/10.1016/j.ijforecast.2008.07.003.

Banternghansa, Chanont, and Michael W. McCracken. 2009. "Forecast Disagreement among FOMC Members." Working paper 2009–059. Federal Reserve Bank of St. Louis. https://fedinprint.org/item/fedlwp/9803/original.

Bao, Yong, Tae-Hwy Lee, and Burak Saltoğlu. 2007. "Comparing density forecast models." *Journal of Forecasting* 26 (3): 203–25. https://doi.org/10.1002/for.1023.

Batchelor, Roy A. 1990. "All Forecasters Are Equal." *Journal of Business & Economic Statistics* 8 (1): 143–44. https://doi.org/10.1080/07350015.1990.10509784.

Batchelor, Roy, and Pami Dua. 1996. "Empirical Measures of Inflation Uncertainty: A Cautionary Note." *Applied Economics* 28 (3): 333–41. https://doi.org/10.1080/000368496328704.

Bates, J. M., and C. W. J. Granger. 1969. "The Combination of Forecasts." *OR* 20 (4): 451–68. https://doi.org/10.2307/3008764.

Beechey, Meredith J., Benjamin K. Johannsen, and Andrew T. Levin. 2011. "Are Long-Run Inflation Expectations Anchored More Firmly in the Euro Area Than in the United States?" *American Economic Journal: Macroeconomics* 3 (2): 104–29. https://doi.org/10.1257/mac.3.2.104.

Berge, Travis J. 2018. "Understanding Survey-Based Inflation Expectations." *International Journal of Forecasting* 34 (4): 788–801. https://doi.org/10.1016/j.ijforecast.2018.07.003.

Berkowitz, Jeremy. 2001. "Testing Density Forecasts with Applications to Risk Management." *Journal of Business and Economic Statistics* 19: 465–74. https://doi.org/10.1198/07350010152596718.

Bernanke, Ben S. 1983. "Irreversibility, Uncertainty, and Cyclical Investment." *The Quarterly Journal of Economics* 98 (1): 85–106. https://doi.org/10.2307/1885568.

Binder, Carola Conces. 2017. "Measuring Uncertainty Based on Rounding: New Method and Application to Inflation Expectations." *Journal of Monetary Economics* 90 (October): 1–12. https://doi.org/10.1016/j.jmoneco.2017.06.001.

Bloom, Nicholas. 2009. "The Impact of Uncertainty Shocks." *Econometrica* 77 (3): 623–85. https://doi.org/10.3982/ecta6248.

Bloom, Nicholas, Stephen Bond, and John Van Reenen. 2007. "Uncertainty and Investment Dynamics." *The Review of Economic Studies* 74 (2): 391–415. https://doi.org/10.1111/j.1467-937X.2007.00426.x.

Bloom, Nicholas, Max Floetotto, Nir Jaimovich, Itay Saporta-Eksten, and Stephen J. Terry. 2018. "Really Uncertain Business Cycles." *Econometrica* 86 (3): 1031–65. https://doi.org/10.3982/ECTA10927.

Boero, Gianna, Jeremy Smith, and Kenneth F. Wallis. 2008a. "Evaluating a Three-Dimensional Panel of Point Forecasts: The Bank of England Survey of External Forecasters." *International Journal of Forecasting* 24 (3): 354–67. https://doi.org/10.1016/j.ijforecast.2008.04.003.

———. 2008b. "Uncertainty and Disagreement in Economic Prediction: The Bank of England Survey of External Forecasters." *The Economic Journal* 118 (530): 1107–27. https://doi.org/10.1111/j.1468-0297.2008.02162.x.

———. 2011. "Scoring Rules and Survey Density Forecasts." *International Journal of Forecasting* 27 (2): 379–93. https://doi.org/10.1016/j.ijforecast.2010.04.003.

———. 2015. "The Measurement and Characteristics of Professional Forecasters' Uncertainty." *Journal of Applied Econometrics* 30 (7): 1029–46. https://doi.org/10.1002/jae.2400.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer. 2020. "Overreaction in Macroeconomic Expectations." *The American Economic Review* 110 (9): 2748–82. https://doi.org/10.1257/aer.20181219.

Bowles, Carlos, Roberta Friz, Veronique Genre, Geoff Kenny, Aidan Meyler, and Tuomas Rautanen. 2007. "The ECB Survey of Professional Forecasters (SPF) – A Review after Eight Years' Experience." Occasional Paper 59. European Central Bank. http://hdl.handle.net/10419/154512.

Brier, Glenn W. 1950. "Verification of Forecasts Expressed in Terms of Probability." *Monthly Weather Review* 78 (1): 1–3. https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.

Broer, Tobias, and Alexandre Kohlhas. 2018. "Forecaster (Mis-)Behavior." CEPR Discussion Paper 12898. C.E.P.R. https://ideas.repec.org/p/cpr/ceprdp/12898.html.

Campbell, John Y. 2000. "Asset Pricing at the Millennium." *The Journal of Finance* 55 (4): 1515–67. https://doi.org/10.1111/0022-1082.00260.

Capistrán, Carlos, and Allan Timmermann. 2009. "Forecast Combination With Entry and Exit of Experts." *Journal of Business & Economic Statistics* 27 (4): 428–40. https://doi.org/10.1198/jbes.2009.07211.

Carroll, Christopher D. 2003. "Macroeconomic Expectations of Households and Professional Forecasters." *The Quarterly Journal of Economics* 118 (1): 269–98. https://doi.org/10.1162/00335530360535207.

Chan, Joshua C. C., Todd E. Clark, and Gary Koop. 2018. "A New Model of Inflation, Trend Inflation, and Long-Run Inflation Expectations." *Journal of Money, Credit and Banking* 50 (1): 5–53. https://doi.org/10.1111/jmcb.12452.

Christensen, Jens H., Francis X. Diebold, Georg H. Strasser, and Glenn D. Rudebusch. 2008. "Multivariate Comparison of Predictive Accuracy." Working paper. Federal Reserve Bank of San Francisco.

Clements, Michael P. 2004. "Evaluating the Bank of England Density Forecasts of Inflation." *The Economic Journal* 114 (498): 844–66. https://doi.org/10.1111/j.1468-0297.2004.00246.x.

———. 2009. "Internal Consistency of Survey Respondents' Forecasts: Evidence Based on the Survey of Professional Forecasters *." In *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, edited by Jennifer Castle and Neil Shephard. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199237197.003.0008.

———. 2010. "Explanations of the Inconsistencies in Survey Respondents' Forecasts." *European Economic Review* 54 (4): 536–49. https://doi.org/10.1016/j.euroecorev.2009.10.003.

———. 2011. "An Empirical Investigation of the Effects of Rounding on the SPF Probabilities of Decline and Output Growth Histograms." *Journal of Money, Credit and Banking* 43 (1): 207–20. https://doi.org/10.1111/j.1538-4616.2010.00371.x.

———. 2014a. "Probability Distributions or Point Predictions? Survey Forecasts of US Output Growth and Inflation." *International Journal of Forecasting* 30 (1): 99–117. https://doi.org/10.1016/j.ijforecast.2013.07.010.

———. 2014b. "Forecast Uncertainty—Ex Ante and Ex Post: U.S. Inflation and Output Growth." *Journal of Business & Economic Statistics* 32 (2): 206–16. https://doi.org/10.1080/07350015.2013.859618.

———. 2015. "Are Professional Macroeconomic Forecasters Able To Do Better Than Forecasting Trends?" *Journal of Money, Credit and Banking* 47 (2–3): 349–82. https://doi.org/10.1111/jmcb.12179.

———. 2018. "Are Macroeconomic Density Forecasts Informative?" *International Journal of Forecasting* 34 (2): 181–98. https://doi.org/10.1016/j.ijforecast.2017.10.004.

———. 2019. "Do Forecasters Target First or Later Releases of National Accounts Data?" *International Journal of Forecasting* 35 (4): 1240–49. https://doi.org/10.1016/j.ijforecast.2018.11.009.

———. 2020. "Are Some Forecasters' Probability Assessments of Macro Variables Better Than Those of Others?" *Econometrics* 8 (2): 1–16. https://doi.org/10.3390/econometrics8020016.

———. 2021a. "Do Survey Joiners and Leavers Differ from Regular Participants? The US SPF GDP Growth and Inflation Forecasts." *International Journal of Forecasting* 37 (2): 634–46. https://doi.org/10.1016/j.ijforecast.2020.08.003.

———. 2021b. "Rounding Behaviour of Professional Macro-Forecasters." *International Journal of Forecasting* 37 (4): 1614–31. https://doi.org/10.1016/j.ijforecast.2021.03.003.

———. 2022. "Forecaster Efficiency, Accuracy, and Disagreement: Evidence Using Individual-Level Survey Data." *Journal of Money, Credit and Banking* 54 (2–3): 537–68. https://doi.org/10.1111/jmcb.12867.

Clements, Michael P., and Ana Beatriz Galvão. 2017a. "Model and Survey Estimates of the Term Structure of US Macroeconomic Uncertainty." *International Journal of Forecasting* 33 (3): 591–604. https://doi.org/10.1016/j.ijforecast.2017.01.004.

———. 2017b. "Predicting Early Data Revisions to U.S. GDP and the Effects of Releases on Equity Markets." *Journal of Business & Economic Statistics* 35 (3): 389–406. https://doi.org/10.1080/07350015.2015.1076726.

———. 2019. "Data Revisions and Real-Time Forecasting." In *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190625979.013.248.

Clements, Michael P., and David I. Harvey. 2009. "Forecast Combination and Encompassing." In *Palgrave Handbook of Econometrics: Volume 2: Applied Econometrics*, edited by Terence C. Mills and Kerry Patterson, 169–98. London: Palgrave Macmillan UK. https://doi.org/10.1057/9780230244405_4.

Coibion, Olivier, and Yuriy Gorodnichenko. 2012. "What Can Survey Forecasts Tell Us about Information Rigidities?" *Journal of Political Economy* 120 (1): 116–59. https://doi.org/10.1086/665662.

———. 2015a. "Is the Phillips Curve Alive and Well after All? Inflation Expectations and the Missing Disinflation." *American Economic Journal: Macroeconomics* 7 (1): 197–232. https://doi.org/10.1257/mac.20130306.

———. 2015b. "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts." *The American Economic Review* 105 (8): 2644–78. https://doi.org/10.1257/aer.20110306.

Coibion, Olivier, Yuriy Gorodnichenko, and Rupal Kamdar. 2018. "The Formation of Expectations, Inflation, and the Phillips Curve." *Journal of Economic Literature* 56 (4): 1447–91. https://doi.org/10.1257/jel.20171300.

Croushore, Dean. 1993. "Introducing: The Survey of Professional Forecasters." *Business Review*, November, 3–15. https://ideas.repec.org/a/fip/fedpbr/y1993inovp3-15.html.

———. 2006. “Chapter 17 - Forecasting with Real-Time Macroeconomic Data.” In *Handbook of Economic Forecasting*, edited by Graham Elliott and Allan Timmermann, 1:961–82. Elsevier. https://doi.org/10.1016/S1574-0706(05)01017-7.

———. 2011a. “Frontiers of Real-Time Data Analysis.” *Journal of Economic Literature* 49 (1): 72–100. https://doi.org/10.1257/jel.49.1.72.

———. 2011b. “Forecasting with Real-Time Data Vintages.” The Oxford Handbook of Economic Forecasting. July 8, 2011. https://doi.org/10.1093/oxfordhb/9780195398649.013.0010.

Croushore, Dean, and Tom Stark. 2019. “Fifty Years of the Survey of Professional Forecasters.” *Economic Insights* 4 (4): 1–11. https://ideas.repec.org/a/fip/fedpei/00042.html.

D’Agostino, Antonello, Kieran Mcquinn, and Karl Whelan. 2012. “Are Some Forecasters Really Better Than Others?” *Journal of Money, Credit and Banking* 44 (4): 715–32. https://doi.org/10.1111/j.1538-4616.2012.00507.x.

D’Amico, Stefania, and Athanasios Orphanides. 2008. “Uncertainty and Disagreement in Economic Forecasting.” Finance and Economics Discussion Series 2008–56. Board of Governors of the Federal Reserve System. https://ideas.repec.org/p/fip/fedgfe/2008-56.html.

———. 2014. “Inflation Uncertainty and Disagreement in Bond Risk Premia.” Working Paper Series WP-2014-24. Federal Reserve Bank of Chicago. https://ideas.repec.org/p/fip/fedhwp/wp-2014-24.html.

Dawid, A. Philip. 1984. “Present Position and Potential Developments: Some Personal Views: Statistical Theory: The Prequential Approach.” *Journal of the Royal Statistical Society: Series A (General)* 147 (2): 278–90. https://doi.org/10.2307/2981683.

Diebold, Francis X., Todd A. Gunther, and Anthony S. Tay. 1998. “Evaluating Density Forecasts with Applications to Financial Risk Management.” *International Economic Review* 39 (4): 863–83. https://doi.org/10.2307/2527342.

Diebold, Francis X., and Roberto S. Mariano. 1995. “Comparing Predictive Accuracy.” *Journal of Business & Economic Statistics* 13 (3): 253–63. https://doi.org/10.2307/1392185.

Diebold, Francis X., Anthony S. Tay, and Kenneth F. Wallis. 1999. “Evaluating Density Forecasts: The Survey of Professional Forecasters.” In *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, edited by Robert F. Engle and Halbert White, 76–90. Oxford, New York: Oxford University Press. https://global.oup.com/academic/product/cointegration-causality-and-forecasting-9780198296836.

Dietrich, J. Kimball, and Douglas H. Joines. 1983. “Rational Expectations, Informational Efficiency, and Tests Using Survey Data: A Comment.” *The Review of Economics and Statistics* 65 (3): 525–29. https://doi.org/10.2307/1924203.

Dixit, Avinash. 1989. “Entry and Exit Decisions under Uncertainty.” *Journal of Political Economy* 97 (3): 620–38. https://doi.org/10.1086/261619.

Dixit, Robert K., and Robert S. Pindyck. 1994. *Investment under Uncertainty*. Princeton University Press. https://doi.org/10.1515/9781400830176.

Dovern, Jonas, and Geoff Kenny. 2020. "Anchoring Inflation Expectations in Unconventional Times: Micro Evidence for the Euro Area." *International Journal of Central Banking* 16 (5): 309–47. https://www.ijcb.org/journal/ijcb20q4a8.htm.

Elliott, Graham, Ivana Komunjer, and Allan Timmermann. 2008. "Biases in Macroeconomic Forecasts: Irrationality or Asymmetric Loss?" *Journal of the European Economic Association* 6 (1): 122–57. https://doi.org/10.1162/JEEA.2008.6.1.122.

Elliott, Graham, and Allan Timmermann. 2004. "Optimal Forecast Combinations under General Loss Functions and Forecast Error Distributions." *Journal of Econometrics* 122 (1): 47–79. https://doi.org/10.1016/j.jeconom.2003.10.019.

Elliott, Graham, Allan Timmermann, and Ivana Komunjer. 2005. "Estimation and Testing of Forecast Rationality under Flexible Loss." *The Review of Economic Studies* 72 (4): 1107–25. https://doi.org/10.1111/0034-6527.00363.

Engelberg, Joseph, Charles F. Manski, and Jared Williams. 2009. "Comparing the Point Predictions and Subjective Probability Distributions of Professional Forecasters." *Journal of Business & Economic Statistics* 27 (1): 30–41. https://doi.org/10.1198/jbes.2009.0003.

———. 2011. "Assessing the Temporal Variation of Macroeconomic Forecasts by a Panel of Changing Composition." *Journal of Applied Econometrics* 26 (7): 1059–78. https://doi.org/10.1002/jae.1206.

Engle, Robert F. 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica* 50 (4): 987–1007. https://doi.org/10.2307/1912773.

Epstein, Edward S. 1969. "A Scoring System for Probability Forecasts of Ranked Categories." *Journal of Applied Meteorology* 8 (6): 985–87. https://doi.org/10.1175/1520-0450(1969)008<0985:ASSFPF>2.0.CO;2.

Faust, Jon, and Jonathan H. Wright. 2009. "Comparing Greenbook and Reduced Form Forecasts Using a Large Realtime Dataset." *Journal of Business & Economic Statistics* 27 (4): 468–79. https://doi.org/10.1198/jbes.2009.07214.

———. 2013. "Chapter 1 - Forecasting Inflation." In *Handbook of Economic Forecasting*, edited by Graham Elliott and Allan Timmermann, 2A:2–56. Elsevier. https://doi.org/10.1016/B978-0-444-53683-9.00001-3.

Favara, Giovanni, and Zheng Song. 2014. "House Price Dynamics with Dispersed Information." *Journal of Economic Theory*, Financial Economics, 149 (January): 350–82. https://doi.org/10.1016/j.jet.2013.05.001.

Figlewski, Stephen, and Paul Wachtel. 1981. "The Formation of Inflationary Expectations." *The Review of Economics and Statistics* 63 (1): 1–10. https://doi.org/10.2307/1924211.

———. 1983. "Rational Expectations, Informational Efficiency, and Tests Using Survey Data: A Reply." *The Review of Economics and Statistics* 65 (3): 529–31. https://doi.org/10.2307/1924204.

Ganics, Gergely, Barbara Rossi, and Tatevik Sekhposyan. 2020. "From Fixed-Event to Fixed-Horizon Density Forecasts: Obtaining Measures of Multi-Horizon Uncertainty from Survey

Density Forecasts." CEPR Discussion Paper 14267. C.E.P.R. https://ideas.repec.org/p/cpr/ceprdp/14267.html.

Garcia, Juan Angel. 2003. "An Introduction to the ECB's Survey of Professional Forecasters." ECB Occasional Paper 8. European Central Bank. http://hdl.handle.net/10419/154461.

Geanakoplos, John. 2010. "The Leverage Cycle." In *NBER Macroeconomics Annual 2009*, edited by Martin S. Eichenbaum, Erik Hurst, and Jonathan A. Parker, 24:1–65. https://doi.org/10.1086/648285.

Genre, Véronique, Geoff Kenny, Aidan Meyler, and Allan Timmermann. 2013. "Combining Expert Forecasts: Can Anything Beat the Simple Average?" *International Journal of Forecasting* 29 (1): 108–21. https://doi.org/10.1016/j.ijforecast.2012.06.004.

Giordani, Paolo, and Paul Söderlind. 2003. "Inflation Forecast Uncertainty." *European Economic Review* 47 (6): 1037–59. https://doi.org/10.1016/S0014-2921(02)00236-2.

———. 2006. "Is There Evidence of Pessimism and Doubt in Subjective Distributions? Implications for the Equity Premium Puzzle." *Journal of Economic Dynamics and Control* 30 (6): 1027–43. https://doi.org/10.1016/j.jedc.2005.05.001.

Glas, Alexander. 2020. "Five Dimensions of the Uncertainty–Disagreement Linkage." *International Journal of Forecasting* 36 (2): 607–27. https://doi.org/10.1016/j.ijforecast.2019.07.010.

Glas, Alexander, and Matthias Hartmann. 2016. "Inflation Uncertainty, Disagreement and Monetary Policy: Evidence from the ECB Survey of Professional Forecasters." *Journal of Empirical Finance*, SI: Euro Zone in Crisis, 39 (December): 215–28. https://doi.org/10.1016/j.jempfin.2016.05.001.

Granger, Clive W. J. 1969. "Prediction with a Generalized Cost of Error Function." *Journal of the Operational Research Society* 20 (2): 199–207. https://doi.org/10.1057/jors.1969.52.

Granger, Clive W. J., and Mark J. Machina. 2006. "Chapter 2 - Forecasting and Decision Theory." In *Handbook of Economic Forecasting*, edited by Graham Elliott, Clive W. J. Granger, and Allan Timmermann, 1:81–98. Elsevier. https://doi.org/10.1016/S1574-0706(05)01002-5.

Granger, Clive W. J., and Ramu Ramanathan. 1984. "Improved Methods of Combining Forecasts." *Journal of Forecasting* 3 (2): 197–204. https://doi.org/10.1002/for.3980030207.

Grothe, Magdalena, and Aidan Meyler. -01. "Inflation Forecasts: Are Market-Based and Survey-Based Measures Informative?" *International Journal of Financial Research* 9 (1): 171–88. https://doi.org/10.5430/ijfr.v9n1p171.

Handley, Kyle, and Nuno Limão. 2015. "Trade and Investment under Policy Uncertainty: Theory and Firm Evidence." *American Economic Journal: Economic Policy* 7 (4): 189–222. https://doi.org/10.1257/pol.20140068.

———. 2017. "Policy Uncertainty, Trade, and Welfare: Theory and Evidence for China and the United States." *The American Economic Review* 107 (9): 2731–83. https://doi.org/10.1257/aer.20141419.

Keane, Michael P., and David E. Runkle. 1990. "Testing the Rationality of Price Forecasts: New Evidence from Panel Data." *The American Economic Review* 80 (4): 714–35.

Kenny, Geoff, Thomas Kostka, and Federico Masera. 2014. “How Informative Are the Subjective Density Forecasts of Macroeconomists?” *Journal of Forecasting* 33 (3): 163–85. https://doi.org/10.1002/for.2281.

———. 2015. “Density Characteristics and Density Forecast Performance: A Panel Analysis.” *Empirical Economics* 48 (3): 1203–31. https://doi.org/10.1007/s00181-014-0815-9.

Knüppel, Malte. 2015. “Evaluating the Calibration of Multi-Step-Ahead Density Forecasts Using Raw Moments.” *Journal of Business & Economic Statistics* 33 (2): 270–81. https://doi.org/10.1080/07350015.2014.948175.

Knüppel, Malte, and Guido Schultefrankenfeld. 2019. “Assessing the Uncertainty in Central Banks’ Inflation Outlooks.” *International Journal of Forecasting* 35 (4): 1748–69. https://doi.org/10.1016/j.ijforecast.2019.03.014.

Knüppel, Malte, and Andreea L. Vladu. 2016. “Approximating Fixed-Horizon Forecasts Using Fixed-Event Forecasts.” Discussion Paper 28/2016. *Discussion Papers*. Deutsche Bundesbank. https://ideas.repec.org/p/zbw/bubdps/282016.html.

Krüger, Fabian. 2017. “Survey-Based Forecast Distributions for Euro Area Growth and Inflation: Ensembles versus Histograms.” *Empirical Economics* 53 (1): 235–46. https://doi.org/10.1007/s00181-017-1228-3.

Lahiri, Kajal, and Fushang Liu. 2006. “Modelling Multi-Period Inflation Uncertainty Using a Panel of Density Forecasts.” *Journal of Applied Econometrics* 21 (8): 1199–1219. https://doi.org/10.1002/jae.880.

Lahiri, Kajal, and Xuguang Sheng. 2010. “Measuring Forecast Uncertainty by Disagreement: The Missing Link.” *Journal of Applied Econometrics* 25 (4): 514–38. https://doi.org/10.1002/jae.1167.

Leahy, John V., and Toni M. Whited. 1996. “The Effect of Uncertainty on Investment: Some Stylized Facts.” *Journal of Money, Credit and Banking* 28 (1): 64–83. https://doi.org/10.2307/2077967.

Liu, Yang, and Xuguang Simon Sheng. 2019. “The Measurement and Transmission of Macroeconomic Uncertainty: Evidence from the U.S. and BRIC Countries.” *International Journal of Forecasting*, Forecasting issues in developing economies, 35 (3): 967–79. https://doi.org/10.1016/j.ijforecast.2019.05.001.

Lorenzoni, Guido. 2009. “A Theory of Demand Shocks.” *The American Economic Review* 99 (5): 2050–84. https://doi.org/10.1257/aer.99.5.2050.

Mankiw, N. Gregory, and Ricardo Reis. 2002. “Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve.” *The Quarterly Journal of Economics* 117 (4): 1295–1328. https://doi.org/10.1162/003355302320935034.

Mankiw, N. Gregory, and Matthew D. Shapiro. 1986. “News or Noise? An Analysis of GNP Revisions.” *Survey of Current Busines* 66 (5): 20–25. https://fraser.stlouisfed.org/title/46/item/10115/toc/87214?start_page=22.

Manski, Charles F. 2008. *Identification for Prediction and Decision*. Harvard University Press. https://doi.org/10.4159/9780674033665.

———. 2011. “Interpreting and Combining Heterogeneous Survey Forecasts.” In *The Oxford Handbook of Economic Forecasting*, edited by Michael P. Clements and David F. Hendry, 457–72. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195398649.013.0017.

Manski, Charles F., and Francesca Molinari. 2010. “Rounding Probabilistic Expectations in Surveys.” *Journal of Business & Economic Statistics* 28 (2): 219–31. https://doi.org/10.1198/jbes.2009.08098.

Manzan, Sebastiano. 2011. “Differential Interpretation in the Survey of Professional Forecasters.” *Journal of Money, Credit and Banking* 43 (5): 993–1017. https://doi.org/10.1111/j.1538-4616.2011.00404.x.

Meyler, Aidan. 2020. “Forecast Performance in the ECB SPF: Ability or Chance?” Working paper 2371. European Central Bank. https://ideas.repec.org/p/ecb/ecbwps/20202371.html.

Mincer, Jacob A., and Victor Zarnowitz. 1969. “The Evaluation of Economic Forecasts.” In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, edited by Jacob A. Mincer, 3–46. Studies in Business Cycles. National Bureau of Economic Research. https://ideas.repec.org/h/nbr/nberch/1214.html.

Morris, Stephen, and Hyun Song Shin. 2002. “Social Value of Public Information.” *The American Economic Review* 92 (5): 1521–34. https://doi.org/10.1257/000282802762024610.

Muth, John F. 1961. “Rational Expectations and the Theory of Price Movements.” *Econometrica* 29 (July): 315–35. https://doi.org/10.2307/1909635.

Nordhaus, William D. 1987. “Forecasting Efficiency: Concepts and Applications.” *The Review of Economics and Statistics* 69 (4): 667–74. https://doi.org/10.2307/1935962.

Novy, Dennis, and Alan M. Taylor. 2020. “Trade and Uncertainty.” *The Review of Economics and Statistics* 102 (4): 749–65. https://doi.org/10.1162/rest_a_00885.

O’Hagan, Anthony, Caitlin E. Buck, Alireza Daneshkhah, J. Richard Eiser, Paul H. Garthwaite, David J. Jenkinson, Jeremy E. Oakley, and Tim Rakow. 2006. *Uncertain Judgements: Eliciting Experts’ Probabilities*. Wiley. https://doi.org/10.1002/0470033312.

Patton, Andrew J., and Allan Timmermann. 2007. “Testing Forecast Optimality Under Unknown Loss.” *Journal of the American Statistical Association* 102 (480): 1172–84. https://doi.org/10.1198/016214506000001176.

Patton, Andrew J., and Allan Timmermann. 2010. “Why Do Forecasters Disagree? Lessons from the Term Structure of Cross-Sectional Dispersion.” *Journal of Monetary Economics* 57 (7): 803–20. https://doi.org/10.1016/j.jmoneco.2010.07.001.

———. 2012. “Forecast Rationality Tests Based on Multi-Horizon Bounds.” *Journal of Business & Economic Statistics* 30 (1): 1–17. https://doi.org/10.1080/07350015.2012.634337.

Pesaran, M. Hashem. 1989. *The Limits to Rational Expectations*. Basil Blackwell.

Poncela, Pilar, and Eva Senra. 2017. “Measuring Uncertainty and Assessing Its Predictive Power in the Euro Area.” *Empirical Economics* 53 (1): 165–82. https://doi.org/10.1007/s00181-016-1181-6.

Rich, Robert W., and Joseph S. 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. https://doi.org/10.1162/rest.2009.11167.

———. 2021a. "A Closer Look at the Behavior of Uncertainty and Disagreement: Micro Evidence from the Euro Area." *Journal of Money, Credit and Banking* 53 (1): 233–53. https://doi.org/10.1111/jmcb.12728.

———. 2021b. "All Forecasters Are Not the Same: Time-Varying Predictive Ability across Forecast Environments." Working paper 21–06. Federal Reserve Bank of Cleveland. https://doi.org/10.26509/frbc-wp-202106.

Rosenblatt, Murray. 1952. "Remarks on a Multivariate Transformation." *The Annals of Mathematical Statistics* 23 (3): 470–72. https://doi.org/10.1214/aoms/1177729394.

Rossi, Barbara, and Tatevik Sekhposyan. 2016. "Forecast Rationality Tests in the Presence of Instabilities, with Applications to Federal Reserve and Survey Forecasts." *Journal of Applied Econometrics* 31 (3): 507–32. https://doi.org/10.1002/jae.2440.

Schaal, Edouard. 2017. "Uncertainty and Unemployment." *Econometrica* 85 (6): 1675–1721. https://doi.org/10.3982/ECTA10557.

Sims, Christopher A. 2003. "Implications of Rational Inattention." *Journal of Monetary Economics*, Swiss National Bank/Study Center Gerzensee Conference on Monetary Policy under Incomplete Information, 50 (3): 665–90. https://doi.org/10.1016/S0304-3932(03)00029-1.

———. 2009. "Inflation Expectations, Uncertainty and Monetary Policy." BIS Working Paper 275. Bank for International Settlements. https://ideas.repec.org/p/bis/biswps/275.html.

Smets, Frank, Anders Warne, and Rafael Wouters. 2014. "Professional Forecasters and Real-Time Forecasting with a DSGE Model." *International Journal of Forecasting* 30 (4): 981–95. https://doi.org/10.1016/j.ijforecast.2014.03.018.

Snowberg, Erik, Justin Wolfers, and Eric Zitzewitz. 2013. "Chapter 11 - Prediction Markets for Economic Forecasting." In *Handbook of Economic Forecasting*, edited by Graham Elliott, Clive W. J. Granger, and Allan Timmermann, 2A:657–87. Elsevier. https://doi.org/10.1016/B978-0-444-53683-9.00011-6.

Söderlind, Paul. 2011. "Inflation Risk Premia and Survey Evidence on Macroeconomic Uncertainty." *International Journal of Central Banking* 7 (2): 113–33. https://www.ijcb.org/journal/ijcb11q2a4.htm.

Stark, Tom. 2013. "SPF Panelists' Forecasting Methods: A Note on the Aggregate Results of a November 2009 Special Survey." Federal Reserve Bank of Philadelphia.

Stekler, Herman O. 1987. "Who Forecasts Better?" *Journal of Business & Economic Statistics* 5 (1): 155–58. https://doi.org/10.1080/07350015.1987.10509571.

Tian, Weidong, and Hong Yan. 2009. "Heterogeneous Beliefs, Imitation, and the Vulnerability of Financial Innovation." Working paper. University of North Carolina at Charlotte.

Tomura, Hajime. 2013. "Heterogeneous Beliefs and Housing-Market Boom-Bust Cycles." *Journal of Economic Dynamics and Control* 37 (4): 735–55. https://doi.org/10.1016/j.jedc.2012.11.002.

Turnovsky, Stephen J., and Michael L. Wachter. 1972. "A Test of the 'Expectations Hypothesis' Using Directly Observed Wage and Price Expectations." *The Review of Economics and Statistics* 54 (1): 47–54. https://doi.org/10.2307/1927494.

Wallis, Kenneth F. 2005. "Combining Density and Interval Forecasts: A Modest Proposal." *Oxford Bulletin of Economics and Statistics* 67 (s1): 983–94. https://doi.org/10.1111/j.1468-0084.2005.00148.x.

Woodford, Michael. 2003. "Imperfect Common Knowledge and the Effects of Monetary Policy." In *Knowledge, Information, and Expectations in Modern Macroeconomics, in Honor of Edmund S. Phelps*, edited by Philippe Aghion, Roman Frydman, Joseph E. Stiglitz, and Michael Woodford, 25–58. Princeton University Press. https://press.princeton.edu/books/ebook/9780691223933/knowledge-information-and-expectations-in-modern-macroeconomics.

Zarnowitz, Victor. 1969. "The New ASA-NBER Survey of Forecasts by Economic Statisticians." *The American Statistician* 23 (1): 12–16. https://doi.org/10.1080/00031305.1969.10481813.

Zarnowitz, Victor, and Louis A. Lambros. 1987. "Consensus and Uncertainty in Economic Prediction." *Journal of Political Economy* 95 (3): 591–621. https://doi.org/10.1086/261473.

Zellner, Arnold. 1986. "Biased Predictors, Rationality and the Evaluation of Forecasts." *Economics Letters* 21 (1): 45–48. https://doi.org/10.1016/0165-1765(86)90119-9.

Appendix Table A1: Surveys of Professional Forecasters

This table provides a list of surveys of professional forecasters along with as much information on features of the surveys that we were able to locate. Some links direct the reader to a website, while in other cases the links direct the reader to a report or a research paper that discusses the survey in some capacity. Some categories may be incomplete.

Bank of England: *Survey of External Forecasters*

- *Background* Started in Feb 1996, quarterly, with 23-38 participants
- *Forecast variables* GDP, Inflation, Unemployment Rate, Policy Rate, Exchange Rate
- *Horizons* 1-, 2-, and 3-year-ahead forecasts
- *Density forecasts* Quarterly for GDP, inflation, unemployment rate, policy rate, and exchange rate at all horizons listed above, histogram
- *Data availability* Micro panel data are available
- *Information from* http://wrap.warwick.ac.uk/1421/, https://academic.oup.com/ej/article/118/530/1107/5088816, and https://onlinelibrary.wiley.com/doi/10.1002/jae.2400

Central Bank of Argentina: *Market Expectations Survey (REM)*

- *Background* Started in Jan 2004, monthly, approximately 34 participants
- *Forecast variables* Inflation, Interest Rate, Nominal Exchange Rate, External Sector: Exports and Imports of Goods or Merchandise, Unemployment Rate, Activity: Growth Rate, Primary Result of the National Non-Financial Public Sector (NFPS)
- *Horizons* 2-month, current & 1-qtr-ahead, current & 1-yr-ahead
- *Density forecasts* Not known
- *Data availability* Aggregated data are available; micro panel data are not disclosed
- *Information from* *http://www.bcra.gob.ar/PublicacionesEstadisticas/Relevamiento_Expectativas_de_Mercado_i.asp and https://www.bis.org/ifc/publ/ifcb30i.pdf*

Central Bank of Brazil: *Market Expectations - Focus Survey*

- *Background* Started in May 1999; data can be submitted daily but respondent must submit at least once a month to be included in published reports; approximately 140 respondents
- *Forecast variables* GDP, Inflation, Short-Term Interest Rates, Exchange Rate, Balance of Payments, Trade Balance, FDI, Net Public Sector Debt

- *Horizons* Inflation over next 12, 13-24 months
- *Density forecasts* Not collected
- *Data availability* Micro data are available for respondents providing consent with a one-year lag; weekly, monthly, and annual reports
- *Information from* https://www.bcb.gov.br/en/monetarypolicy/marketexpectations and FAQ – Market Expectations System (bcb.gov.br)

Central Bank of Iceland: *Survey of Market Expectations*

- *Background* Started in 2012, quarterly, approximately 31 respondents
- *Forecast variables* Inflation, Policy Rate, Exchange Rate
- *Horizons* Current, 1-qtr & 2-qtr-ahead, inflation—2- & 10-year-ahead
- *Density forecasts* Not known
- *Data availability* Micro panel data are not available
- *Information from* *https://www.cb.is/statistics/various-measures-of-inflation-expectations/survey-of-market-expectations/*

Central Bank of Indonesia: *Consensus Forecasts*

- *Background* Started in 2001, quarterly, approximately 200 economists
- *Forecast variables* GDP, Inflation, Exchange Rates
- *Horizons* Average for current and 1-yr-ahead, 4-qtr average
- *Density forecasts* Not known
- *Data availability* Micro panel data availability is not known
- *Information from* *https://www.bis.org/ifc/events/7ifcconf_wuryandani_mardiani.pdf*

Central Bank of Israel: *Macroeconomic Staff Forecast*

- *Background* Started in 2001, daily (at least for some variables), approximately 14 respondents
- *Forecast variables* Inflation, Bank of Israel Policy Rate, Exchange Rates; GDP Added in 2017
- *Horizons* Current, 1-, 2- & 12-month-ahead; current & 1-year-ahead
- *Density forecasts* Not collected
- *Data availability* Micro panel data are available upon request
- *Information from* *https://www.sciencedirect.com/science/article/pii/S1059056021001350#*

Central Bank of Turkey: *Survey of Market Participants*

- *Background* Started in 2001, monthly since 2013, approximately 57 respondents
- *Forecast variables* Inflation, Exchange Rates, Current Account Balance, GDP, Repo & Reverse Repo Rates
- *Horizons* Inflation--Current, 1-, 2-month-ahead; end of current, 1- & 2-yr-ahead current account—current & 1-yr-ahead GDP—current & 1-yr-ahead average; Repo rates—end of current month
- *Density forecasts* 12-month-ahead annual consumer inflation expectations and 24-month-ahead annual consumer inflation expectations
- *Data availability* Micro panel data are not available
- *Information from* Email correspondence and *https://www.tcmb.gov.tr/wps/wcm/connect/EN/TCMB+EN/Main+Menu/Statistics/Tendency+Surveys/Survey+of+Market+Participants/*

Conference Board of Canada: *Survey of Forecasters*

- *Background* Started in 1999 Q4, quarterly, 500 firms
- *Forecast variables* GDP, Consumption Spending, and the Current Account Balance
- *Horizons* 2 year
- *Density forecasts* Not collected
- *Data availability* Aggregate data available with subscription; micro panel data are not available
- Information from Email correspondence and https://www.bankofcanada.ca/wp-content/uploads/2010/06/cunningham.pdf

Consensus Economics: *Consensus Forecasts*

- *Background* Started in 1989, monthly, approximately 700 economists
- *Forecast variables* GDP, Inflation, Short-Term Interest Rates, Exchange Rate
- *Horizons* Vary by forecast variable
- *Density forecasts* No
- *Data availability* Micro panel data are not available; consensus estimates are available on a subscription basis
- *Information from* *https://www.consensuseconomics.com/*

Dutch National Bank: *Inflation Expectations Survey*

- *Background* Started in July 2010, weekly with approximately 25 respondents
- *Forecast variables* Euro area inflation
- *Horizons* Current year, next year, and 10-yr-ahead
- *Density forecasts* Quarterly for short- & long-term inflation expectations, histogram
- *Data availability* Aggregate data available; micro panel data are not available
- *Information from* *https://www.bis.org/publ/work809.pdf*

European Central Bank: *Survey of Professional Forecasters*

- *Background* Started in January 1999, quarterly, approximately 60 respondents
- *Forecast variables* GDP, Inflation, and Unemployment Rate, with Non-Compulsory Section on Wage Growth, Price of Oil, and Exchange Rate
- *Horizons* Current calendar year, next calendar year, the calendar year after that, 1-year-ahead, 1-year/1-year forward, and a longer horizon
- *Density forecasts* Quarterly for inflation, core inflation, GDP growth, and unemployment rate at all horizons listed above, histogram
- *Data availability* Micro panel data are available
- *Information from* https://www.ecb.europa.eu/stats/ecb_surveys/survey_of_professional_forecasters/html/index.en.html

Federal Reserve Bank of Philadelphia: *Survey of Professional Forecasters*

- *Background* Started in Q4 1968, quarterly, approximately 40 respondents
- *Forecast variables* GDP, Inflation, Unemployment Rate, Various Business Indicators, and Components of Real GDP
- *Horizons* 1-5-quarter-ahead levels and growth rates, Q4/Q4 growth, year-over-year growth, and a longer horizon
- *Density forecasts* Quarterly for real GDP growth, inflation, and unemployment rate at varying short- and medium-run horizons
- *Data availability* Micro panel data are available
- *Information from* https://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters

Federal Reserve Bank of Philadelphia: *Livingston Survey*

- *Background* Started in June 1946, semi-annual (June & Dec), 10-20 respondents

- *Forecast variables* Quarterly: GDP, Non-Residential FI, After-Tax Corporate Profits; Monthly: IP, Housing Starts, PPI, CPI, Unemployment Rate, Avg Weekly Earnings in Mfg., Retail Sales, Auto Sales, Prime Interest Rate; 3-Month T-Bill Rate, 10-Year T-Note Rate, S&P 500
- *Forecast horizons* Quarterly: current, 1-, 2-, & 4-qtrs-ahead, June—current & 1-yr-ahead, Dec—current yr, 1- & 2-yr-ahead; Monthly: current, 6- & 12-mth, June—current & 1-yr-ahead, Dec—current, 1- & 2-yr-ahead; June 1990—10-yr GDP & inflation
- *Density forecasts* No
- *Data availability* Micro panel data are available, and summaries are published semi-annually
- *Information from* Livingston Survey (philadelphiafed.org)

Japan Center for Economic Research: *ESPF Survey*

- *Background* Started in April 2004, monthly, approximately 40 respondents
- *Forecast variables* GDP, Inflation Less Fresh Food, Unemployment Rate
- *Horizons* GDP – quarter to quarter changes for current, 1-, 2- & 3-quarter-ahead; Inflation – year-on-year change
- *Density forecasts* Not known
- *Data availability* Aggregated data are available
- *Information from* https://www.jcer.or.jp/en/esp-forecast-top and dp2017-04.pdf (kanagawa-u.ac.jp)

Ministry of Finance of the Czech Republic: *Survey of Macroeconomic Forecasts*

- *Background* Started 1996, 2x/year, approximately 14 institutions
- *Forecast variables* GDP, Inflation, Oil Price, Exchange Rates, Interest Rates, Employment, Unemployment Rate, Wages
- *Horizons* 3 years
- *Density forecasts* Not known
- *Data availability* Not known
- *Information from* https://www.mfcr.cz/en/statistics/survey-of-macroeconomic-forecasts

Monetary Authority of Singapore: *Survey of Professional Forecasters*

- *Background* Started in 1999 Q4, quarterly, approximately 20-30 respondents
- *Forecast variables* GDP and Components, Headline and Core Inflation, Unemployment Rate, Short-Term Interest Rates, Exchange Rate
- *Horizons* Annual average for current and next year, rolling horizon for quarters left in current year

- *Density forecasts* GPP in 2001 Q3, inflation in 2017 Q1; histogram
- *Data availability* Micro panel data are available
- *Information from* *Survey of Professional Forecasters Data Documentation (mas.gov.sg)*

National Association for Business Economics: *NABE Outlook*

- *Background* Started in 1965, quarterly with approximately 50 respondents
- *Forecast variables* Macro variables
- *Horizons* Not known
- *Density forecasts* Not collected
- *Data availability* Micro panel data are not available; consensus estimates are available on a subscription basis
- *Information from* https://www.nabe.com/NABE/Surveys/Outlook_Surveys/december-2021-Outlook-Survey-Summary.aspx

National Bank of Poland: *Survey of Professional Forecasters*

- *Background* Started in 2001, quarterly, approximately 20-30 respondents
- *Forecast variables* GDP, Inflation, Unemployment Rate Exchange Rate, Oil Prices, Average Wage Growth, Euro Area GDP
- *Horizons* Point forecasts – current year, 1-year, 2-year; Density forecasts – current year, 1-year, 2-year, annual average over next 5 years
- *Density forecasts* Cumulative density – 5%, median, 95%
- *Data availability* Micro panel data are available
- *Information from* https://amakro.nbp.pl/amakro-forecaster/pages/about.nbp

National Bank of Ukraine: Survey of Experts

- *Background* Started in July 2014, 6x/year, approximately 17 respondents
- *Forecast variables* Inflation, Exchange Rate
- *Horizons* 1-year and 3-year
- *Density forecasts* No
- *Data availability* Aggregated and micro panel data are available
- *Information from* https://voxukraine.org/en/inflation-expectations-in-ukraine-a-long-path-to-anchoring-en/ and https://bank.gov.ua/en/statistic/nbusurvey

Wolters Kluwer: *Blue Chip Economic Indicators*

- *Background* Started in 1976, monthly, approximately 50 respondents
- *Forecast variables* GDP, Inflation, Unemployment Rate, IP, Disposable Income, Non-Residential Fixed Investment, Pre-Tax Corporate Profits, 3-Month T-Bill Rate, Housing Starts, Auto and Light Truck Sales, Net Exports
- *Horizons* Inflation 0 to 7 quarters ahead for the United States, 1 to 2 years ahead for other major economies
- *Density forecasts* Not collected
- *Data availability* Micro panel data available with subscription
- *Information from* https://www.wolterskluwer.com/en/solutions/vitallaw-law-firms/blue-chip

Reserve Bank of Australia: *Survey of Market Economists*

- *Background* Started in 1993, quarterly, approximately 10-20 respondents
- *Forecast variables* Inflation
- *Horizons* Year-end for current year and next year are asked in June and December, 5–10-year average added in mid-2015
- *Density forecasts* No
- *Data availability* Micro panel data are available
- Information from https://www.rba.gov.au/publications/rdp/2016/pdf/rdp2016-02.pdf

Reserve Bank of India: *Survey of Professional Forecasters*

- *Background* Started in Sept 2007, quarterly, approximately 40 respondents
- *Forecast variables* GDP, Inflation, Consumption, Fixed Capital Formation, Imports, Exports, Current Account Deficit, Gross Value Added
- *Horizons* Annual and quarterly averages
- *Density forecasts* For GDP and inflation
- *Data availability* Micro panel data are not available
- *Information from* https://rbidocs.rbi.org.in/rdocs/Publications/PDFs/SPF06022020E68670BBBFD5479B8B5785AD4F85FB43.PDF

Reserve Bank of New Zealand: *Survey of Expectations*

- *Background* Started in 1987, quarterly, approximately 179 respondents
- *Forecast variables* Inflation, Policy Rate, 10-Yr Gov't Bond Rate, GDP, Annual Wage Growth, Unemployment Rate, Exchange Rate, House Price

Index

- *Horizons* Inflation—1-, 2-, 5-yr-ahead; policy rate—current qtr, 1-yr-ahead; 10-yr gov't bond rate—end qtr, 1-yr-ahead; GDP, annual wage growth, unemployment rate, house price index—1- & 2-yr-ahead
- *Density forecasts* Not known
- *Data availability* Micro panel data are not available
- *Information from* https://www.rbnz.govt.nz/statistics/m14

South African Reserve Bank: *Inflation Expectations Survey*

- *Background* Started in Sept 2000, quarterly, approximately 40 financial sector respondents
- *Forecast variables* Headline & Core Inflation, GDP, Prime Interest Rate, Exchange Rate, Gov't Bond Rate, M3 Supply Growth, Capacity Utilization in Mfg., Wage and Salary Growth
- *Horizons* Current, 1- & 2-year-ahead; wage & salary – current & 1-yr-ahead
- *Density forecasts* Expected inflation
- *Data availability* Aggregated data are available; access to micro panel data is limited
- *Information from* https://www.ber.ac.za/Reasearch/Method/Inflation-expectations/#IEShow

ZEW: *Financial Market Survey*

- *Background* Started in 1991, monthly, approximately 270 (German) respondents
- *Forecast variables* GDP, Inflation, Exchange Rates, Oil Prices for Eurozone, Germany, Japan, U.S., U.K., France, and Italy
- *Horizons* Year-end for current year and next year are asked in June and December, 5–10-year average added in mid-2015
- *Density forecasts* No
- *Data availability* Micro panel data are not available
- *Information from* https://ftp.zew.de/pub/zew-docs/div/Kurzinfo_English.pdf and https://www.zew.de/en/publications/zew-expertises-research-reports/research-reports/business-cycle/zew-financial-market-survey