Inference on Probabilistic Surveys in Macroeconomics with an Application to the Evolution of Uncertainty in the Survey of Professional Forecasters during the COVID Pandemic

Federico Bassetti
Polytechnic University of Milan

Roberto Casarin
Ca' Foscari University of Venice

Marco Del Negro*
Federal Reserve Bank of New York

February 3, 2022

Abstract

Probabilistic surveys on macroeconomic variables provide a wealth of information to the applied researcher. Extracting and using this information is not a trivial task, however. This chapter discusses the challenges involved in this task and the approaches used so far in the literature for conducting inference on probabilistic surveys. It also provides an application of some of these methods using the U.S. Survey of Professional Forecasters and investigates the evolution of uncertainty and tail risk for both output growth and inflation during the COVID pandemic.

JEL CLASSIFICATION: C1, C11, C13, C15, C32, C58, G12, G13, G15

KEY WORDS: Surveys, Probabilistic Forecasts, Uncertainty, Tail Risks, Expert Opinion Pools, Bayesian Non-parametric

*Correspondence: Marco Del Negro (marco.delnegro@ny.frb.org): Research Department, Federal Reserve Bank of New York, 33 Liberty Street, New York NY 10045. Other correspondence: Roberto Casarin (r.casarin@unive.it) and Federico Bassetti (federico.bassetti@polimi.it). The views expressed in this paper do not necessarily reflect those of the Federal Reserve Bank of New York or the Federal Reserve System.

I Introduction

Manski (2004) made economists appreciate the advantages of probabilistic surveys relative to surveys that simply ask respondents for their point projections. Probabilistic surveys provide information about the entire predictive distribution, which can in principle be used to extract subjective measures of mean predictions, uncertainty, and tail risks. Partly following Manski (2004)'s lead, and partly because of the growing interest in studying uncertainty in the aftermath of the Great Recession, a large number of probabilistic surveys in macroeconomics have emerged over the last decade or so, which elicit predictive probabilities on macroeconomic variables from professional forecasters, financial market participants, consumers, and firms, in the US and other countries. Examples are the New York Fed's Survey of Consumer Expectations (see Armantier et al., 2017), the New York Fed's Survey of Primary Dealers and Market Participants, the Atlanta Fed's Survey of Business Uncertainty and Business Inflation Expectations Survey (see Altig et al., 2020), the European Central Bank's Survey of Professional Forecasters and the Bank of England's Survey of External Forecasters (see Boero et al., 2008a), in addition to the already established Philadelphia Fed's U.S. Survey of Professional Forecasters (henceforth, SPF).

Yet, respondents to probabilistic surveys do not generally provide macroeconomists with direct information on the objects they are interested in, such as various moments (e.g., means, variances, skewness) and quantiles of their forecast distribution. Instead, they provide probabilities associated with generally pre-specified (by the survey designer) bins. Extracting and using the information provided by respondents is not a trivial task, however. This chapter discusses the challenges involved in this task and the approaches used so far in the literature for conducting inference on probabilistic surveys. While the focus of the chapter is on the inference problem, it also discusses some of the literature that made use of probabilistic surveys in macroeconomics—although we keep this review to a minimum given that some of this literature is already covered in ClementsetalCHAPT-inVOLUME (2022).

Finally, the chapter also provides an application of some of these inference methods using the SPF. In particular, it investigates the evolution of uncertainty and tail risk for both output growth and inflation during the COVID pandemic.

The structure of the chapter is as follows. Section II discusses the approaches used in conducting inference on probabilistic surveys and some related econometric issues that arise. In particular, Section II.A describes the inference problem posed by probabilistic forecasts, while Section II.B reviews existing approaches and their limitations. Section II.C describes

a Bayesian non-parametric approach based on Del Negro et al. (2018), which tries to address some of the issues mentioned in Section II.B, with a particular emphasis on inference for measures of uncertainty and tail risk. This Section also includes a comparisons between the different approaches in a real data application. Section III discusses some of the challenges in assessing the uncertainty of forecasters, while Section IV documents the heterogeneity in density predictions across forecasters, again focusing on uncertainty. Section V gives an overview of some existing approaches for aggregating individual responses (opinion pools). Finally, the Bayesian non-parametric method discussed in Section II.C is used in Section VI to analyze the evolution of US SPF density forecasts for GDP growth and inflation during the COVID pandemic. The analysis focuses on how the COVID pandemic affected the consensus forecast, average measures of subjective uncertainty, as well as heterogeneity in mean projections, uncertainty, and tail probabilities.

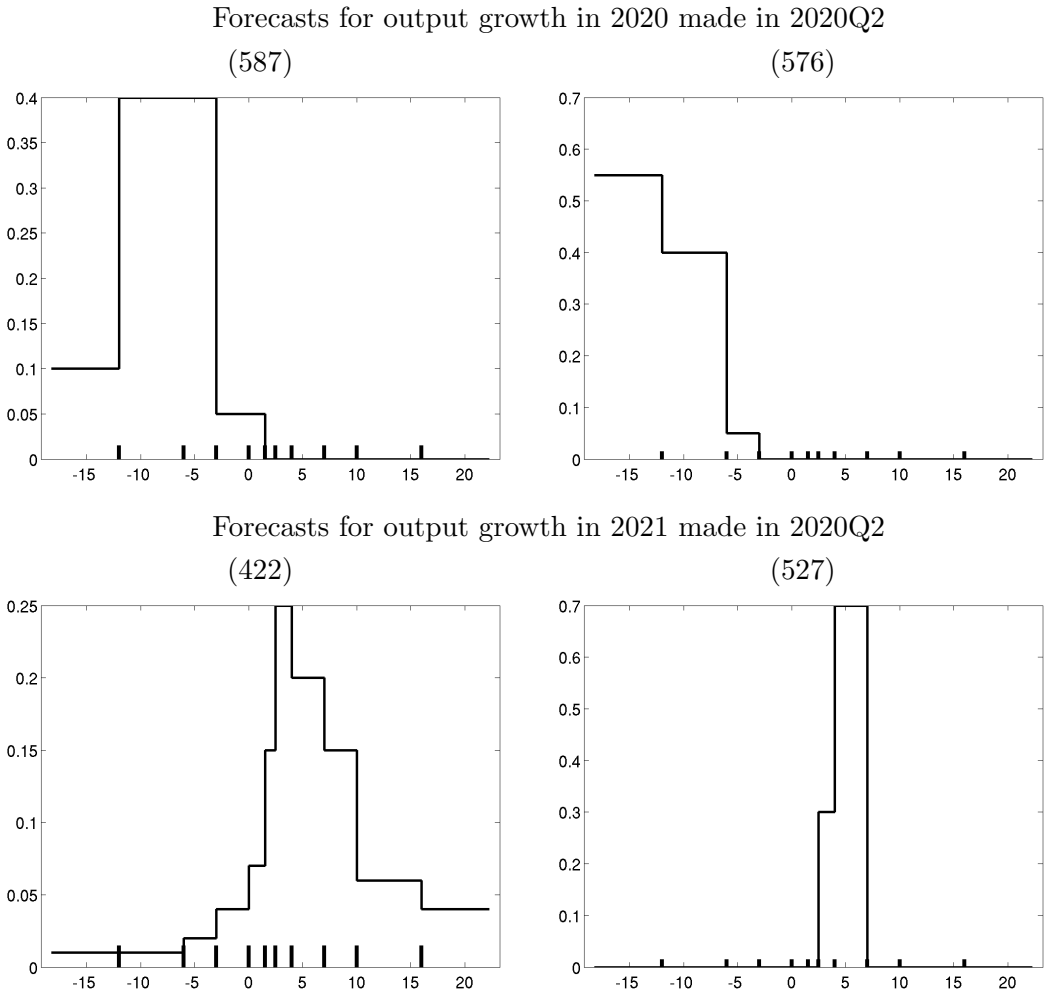
II Inference on Probabilistic Surveys

This section reviews current approaches used for inference on probabilistic surveys. It discusses some challenges these approaches face, while attempting to address some of them based on recent work by Del Negro et al. (2018).

II.A The Inference Problem

Probabilistic forecasts take the form of probabilities assigned to bins: the percent chance that the variable of interest, such as inflation or GDP growth, falls within different contiguous ranges. In most of the surveys, the bins are pre-specified by the survey designer (e.g., the Philadelphia Fed's SPF), although some recent surveys only specify the number of bins and let the respondents determine their boundaries (e.g., the Atlanta Fed's Survey of Business Uncertainty, where a 5-point probability is provided by each survey respondent). The bins are mutually exclusive and contiguous, and generally cover the entire real line whenever the variable being forecast, y, is continuous and can take values in the interval $(-\infty, +\infty)$. In what follows, $(y_{j-1}, y_j]$, $j = 1, \ldots, J$ will be a set of bins such that $y_0 < y_1 < \ldots < y_J$, where y_0 and y_J are equal to $-\infty$ (left open bin) and $+\infty$ (right open bin), respectively. For each forecaster $i = 1, ..., n$, the available data consists of a vector of probabilities $\mathbf{z}_i = (z_{i,1}, \ldots, z_{i,J})$, with $z_{i,j} \geq 0$ and $z_{i,1} + z_{i,2} + \ldots + z_{i,J} = 1$, measuring the predictive likelihood that y falls within the respective bins.

Figure 1: Probability Forecasts for Selected Examples from the 2020Q2 SPF



Note: Each panel displays the forecast probabilities $z_{i,j}$, $j = 1, \ldots, 11$ (step-wise solid lines) for a given forecaster (forecaster number shown in parentheses) and the bin bounds (black ticks, horizontal axis).

Figure 1 shows the probabilities that the 2020 (first row) and the 2021 (second row) real GDP growth falls within the bins (horizontal axis) as provided by four respondents in the US SPF conducted in the second quarter of 2020, specifically in mid-May (from now on, we will use the notation XQY to denote the survey made in quarter Y of year X). These probabilities are displayed as histograms, while the black ticks on the horizontal axis mark the boundaries of the bins. In 2020Q2 the survey histogram has 11 bins with bounds -12, -6, -3, 0, 1.5, 2.5, 4, 7, 10, and 16, with left and right open bins. In the SPF the bins have been changing over time. Stark (2013) discusses at length some of the features of the SPF survey, and the Philadelphia Fed's site provides a manual for interpreting the data that includes the history up to the present of bin boundaries for the various variables being forecast. The fact that

the bin boundaries change over time needs to be borne in mind when comparing surveys for different years.

These four examples illustrate a few common features of the SPF data: potential asymmetry, the presence of probability mass on open bins (forecaster 576 assigns most probability mass to the left open bin), the fact that some forecasters assign positive probability to most if not all bins (forecaster 422) while for others most bins have zero probability (forecaster 527), and rounding (almost all probabilities in Figure 1 are round numbers, with some bins for forecaster 422 being the only exception). The econometrician's problem is to use those few points given by the elements of the survey probability vector $\mathbf{z}_i$ of the i-th forecaster to address a number of questions of interest: What is the mean prediction for forecaster i? How uncertain are they? What probability do they place on tail events, e.g., output growth in 2020 or 2021 being below 10 percent? The next section describes how the literature has gone after this problem, considering the survey features mentioned above.

II.B Current Approaches

This section describes the approaches used so far for translating the information provided by the respondents into the objects of interest mentioned above. The general approach for macroeconomic surveys has been to postulate that forecasters $i = 1, \ldots, n$ have in mind a given predictive probability distribution $F_i(y)$ over the variable being forecast, which they use to assign the bin probabilities $\mathbf{z}_i$. The name of the game for the econometrician is then to conduct inference on $F_i(y)$ based on the data $\mathbf{z}_i$, and then use the estimated $F_i(y)$ to answer the questions of interest.

To our knowledge, all existing literature has accomplished this task by fitting a given *parametric distribution* to the Cumulative Distribution Function (CDF) implied by the bin probabilities, respondent by respondent, that is fitting $Z_{ij} = z_{i,1} + \cdots + z_{i,j}$ $j = 1, \ldots, J$, $i = 1, \ldots, n$ using a parametric family of distributions $\{F(y|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$. The type of the parametric distribution varies across studies, from a mixture of uniforms/piece-wise linear CDF (that is, assuming that the probability is uniformly distributed within each bin; Zarnowitz and Lambros, 1987), to a Gaussian (Giordani and Soderlind, 2003), a skew-normal (Garcia and Manzanares, 2007), a generalized beta (Engelberg et al., 2009)[1] and a skew-t distribution (e.g., Ganics et al., 2020). The Gaussian and the generalized beta

[1] Whenever the number of (adjacent) bins with positive probability is two or fewer, Engelberg et al. (2009) uses a triangular distribution.

assumptions have been the most popular approaches in academic research, as discussed in ClementsetalCHAPT-inVOLUME (2022), although in applied work at central banks the mixture of uniforms approach is often followed. The parameters of each distribution are usually estimated using nonlinear least squares, respondent by respondent; that is, $F_i(y) = F(y|\boldsymbol{\theta}_i^*)$, where

$$\boldsymbol{\theta}_i^* = \underset{\boldsymbol{\theta}_i}{\operatorname{argmin}} \sum_{j=1}^{J} \left| Z_{ij} - F(y_j|\boldsymbol{\theta}_i) \right|^2. \tag{1}$$

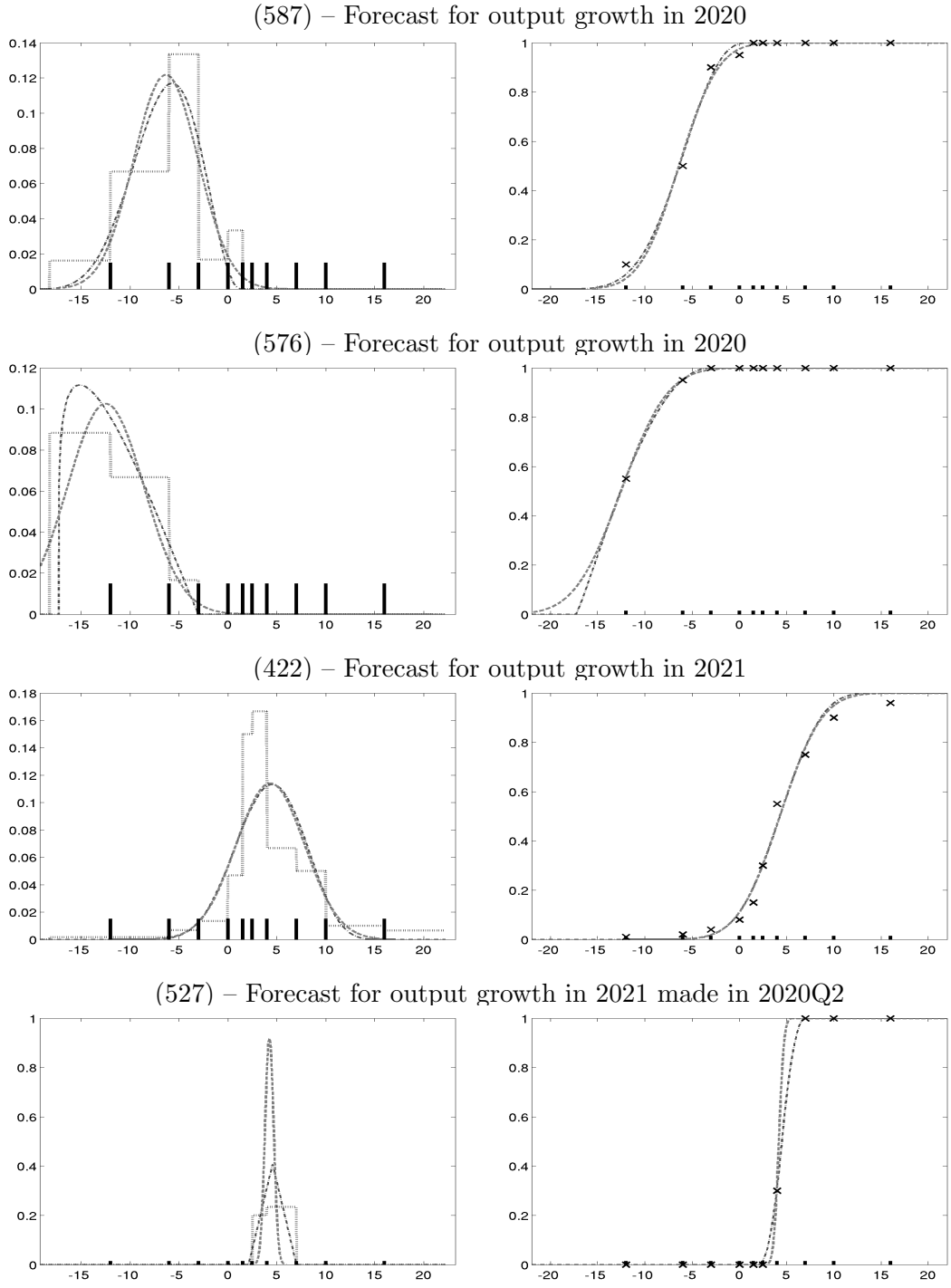
These approaches have been popular but have some limitations, which are generally well known in the literature and which we discuss in the remainder of this section using the examples of Figure 1, focusing on the generalized beta and Gaussian cases.[2] A first limitation is that the assumed parametric distribution may be misspecified, in the sense that it may not fit the individual responses very well. For instance, the right column panels of Figure 2 show that for respondents 587 and 422 neither the normal nor the beta distributions fit all the observed Z_{ij}'s. Second, the width of the bins can be large, as is the case for real output growth in 2020Q2, implying that even if the distributions fit the Z_{ij}'s well, the inference results on moments and quantiles can be sensitive to the distributional assumption.

Another issue is that bounded distributions such as the beta or the mixture of uniforms take literally the z_{ij} that are zero, in that they place no probability mass on bins where the respondents place no mass (see respondents 587, 576, and 527). Since y is a continuous variable, one suspects that $z_{ij} = 0$ may not literally mean that $F_i(y)$ places no mass on that interval, but that the mass is relatively small so that the respondent felt comfortable reporting zero. More generally, for all assumed $F(\cdot)$'s the approach outlined in expression (1) ignores the issue of rounding, in that it takes all the Z_{ij}'s literally even though, as mentioned before, they are all round numbers suggesting that the respondent reported approximate probabilities (Dominitz and Manski, 1996; D'Amico and Orphanides, 2008; Boero et al., 2008b, 2014; Engelberg et al., 2009; Manski and Molinari, 2010; Manski, 2011; Giustinelli et al., 2020, among others, discuss the issue of rounding; Binder, 2017 uses rounding to measure uncertainty).[3]

[2] In what follows, when showing results using the beta assumption we follow Engelberg et al. (2009) quite literally. We also followed the literature in terms of closing the open bins for both the generalized beta and the step-wise uniform.

[3] Manski and Molinari (2010) and Giustinelli et al. (2020) propose to treat the issue of rounding by considering interval data and using a person's response pattern across different questions to infer her or his rounding practice. It is important to note that the inferential approach based on interval data followed by these researchers is very different from the one described at the beginning of this section.

Figure 2: Inference Using the Uniform, Beta, and Gaussian Distributions: PDFs and CDFs for Selected Examples from the 2020Q2 SPF



Note: Nonlinear least squares estimation of the subjective PDF (left) and CDF (right) using the normal (gray, dashed line) or beta (black, dash-and-dotted line) parametric assumptions. In addition, the left column shows the step-wise uniform PDF (dotted lines) obtained from step-wise uniform PDF (dotted lines) implied by the histogram probabilities z_{ij}, $j = 1, \ldots, J$. The right column displays the observed cumulated histogram probabilities Z_{ij} $j = 1, \ldots, J$ (crosses).

Finally, and very importantly, existing approaches do not capture inference uncertainty. Because of the limited information available, the econometrician cannot be sure about the inferred CDF $F_i(\cdot)$. Yet, most if not all existing approaches completely ignore this inference uncertainty, even that concerning $\boldsymbol{\theta}_i$ for a given parametric assumption, let alone the uncertainty about the shape of $F_i(\cdot)$. For respondent 527 in Figure 2 there are arguably many Gaussians with a fit very close to that of the one implied by $\boldsymbol{\theta}_i^*$.[4] This omission implies that confidence bands and hypothesis testing procedures cannot be derived.

As mentioned, these limitations are well known in the literature. There have been attempts to address some of these issues, in particular the potential misspecification, by choosing more flexible families of distributions such as the skew-normal or the skew-Student-t distribution (e.g., Garcia and Manzanares, 2007; Ganics et al., 2020). But the risk of misspecification obviously remains; for example, some forecasters display multimodal histograms (see Del Negro et al., 2018). Most importantly, if the econometrician does not account for inference uncertainty, this flexibility comes at the price of possible overparamterization. Again, respondent 527 has a histogram with only two bins, and one can hardly discriminate between normal and Student-t distributions. The next section discusses an inference method that attempts to overcome some of the limitations of the current approaches.

II.C A Bayesian Non-parametric Alternative

The Bayesian method developed in Del Negro et al. (2018) relies on a probabilistic model for the forecaster's subjective probabilities and a flexible (non-parametric) modeling approach. We will refer to this method as BNP. It differs from existing methods in four important dimensions. First, it is robust to misspecification regarding the parametric assumption for predictive CDF $F_i(\cdot)$. Second, the method allows for full-fledged inference regarding the mapping between data and objects of interest (e.g., the quantiles of the predictive density), thereby providing posterior probabilities that reflect inference uncertainty. Third, it conducts inference *jointly* across survey respondents, that is, using the entire cross-section

[4] Researchers recognize the emergence of an inference issue especially when the information provided by the respondent is very limited, but the proposed solution mostly amounts to either choosing less parameterized distributions or discarding the respondent. For instance, some researchers simply discard histograms with fewer than three bins Clements (2010), others (Engelberg et al., 2009; Clements, 2014b,a; Clements and Galvão, 2017) use a triangle distribution in these cases, as mentioned above. Liu and Sheng (2019), however, make an attempt to account for parameter uncertainty for given parametric assumptions. They propose maximum likelihood estimation of parametric distributions on artificial data generated from the histogram.

instead of being applied to each respondent separately. The joint inference allows for partial information pooling across forecasters and exploits commonalities across forecasters to improve our inference for the aggregate and individual CDFs. This implies that when the number of forecasters grows posterior estimates become more precise, making it possible to obtain some consistency results, as discussed later. Fourth, it explicitly accounts for noise (including rounding toward zero) in the survey responses.

This section offers a mostly verbal description of the approach, keeping the analytical expressions to a minimum. Del Negro et al. (2018) provides any missing detail and contains formal derivations of the results mentioned in this section. The model description works as follows. We first describe a *parametric* model. This model follows the literature in that it assumes that forecasters have in mind a specific predictive distribution $F(\cdot)$ which they use to assign probabilities $\boldsymbol{\nu}$ to the bins. It is different from the literature in that it explicitly postulates that the data $\mathbf{z}$ are noise-ridden versions of the $\boldsymbol{\nu}$'s, where again the noise is assumed to have some parametric form. We then depart from this parametric framework by embedding it into a more general Bayesian *non-parametric* approach, thereby amending the potential misspecification associated with the parametric assumptions.

A parametric probabilistic model

The probabilities $\mathbf{z}_i$ reported by the forecaster are imperfect representations of the respondent forecast uncertainty (see Boero et al., 2008a and the discussion in ClementsetalCHAPTinVOLUME, 2022), and can be considered noise-ridden measurements of an unobserved vector of subjective probabilities over the J bins $\boldsymbol{\nu}_i = (\nu_{i1}, \ldots, \nu_{iJ})$ with $\nu_{ij} \geq 0$ and $\nu_{i1} + \ldots + \nu_{iJ} = 1$. The uncertainty in the vector of probabilities $\mathbf{z}_i$ is encoded into a probability distribution $h(\cdot)$,

$$\mathbf{z}_i = (z_{i,1}, \ldots, z_{i,J}) \sim h(\mathbf{z}_i | \boldsymbol{\nu}_i; \boldsymbol{\theta}_i), \tag{2}$$

which captures the noise due to approximations or to actual mistakes in reporting. In choosing $h(\cdot)$, one needs to account for the fact that $\mathbf{z}_i$ belongs to the simplex; that is, the elements of $\mathbf{z}_i$ are positive and sum up to one. A natural choice for random variables on the simplex is the Dirichlet distribution. A drawback of the standard Dirichlet distribution is that its PDF is null for $\mathbf{z}_i$'s that have some elements equal to zero, when in fact forecasters often assign zero probability to one or more bins. Del Negro et al. (2018) therefore use the Zadora distribution (see Zadora et al. (2010); Scealy and Welsh (2011)), which allows for values of the random vector that are zeros.

The vector of probabilities over the J bins can be re-parameterized with a family of subjective probability functions $F(y|\boldsymbol{\theta})$:

$$\nu_{ij}(\boldsymbol{\theta}_i) = F(y_j|\boldsymbol{\theta}_i) - F(y_{j-1}|\boldsymbol{\theta}_i),\ j = 1, \ldots, J. \tag{3}$$

This leads us to write $h(\mathbf{z}_i|\boldsymbol{\nu}_i; \boldsymbol{\theta}_i)$ as $h(\mathbf{z}_i|\boldsymbol{\theta}_i)$, where $\boldsymbol{\theta}_i \in \Theta$ includes both the parameters describing the CDF $F(\cdot)$ and those needed to specify $h(\cdot)$ in (2). In the application below, we follow Del Negro et al. (2018) and assume $F(\cdot)$ is a mixture of two normal distributions.[5]

A Bayesian non-parametric approach

The Bayesian non-parametric approach amounts to a flexible hierarchical setup where each respondent is described by a potentially infinite mixture of the parametric probability distribution described so far. These mixtures can be seen as forecaster "types" (for concreteness, let us think of low and high uncertainty; low and high mean; left-skewed and right-skewed; low and high reporting noise; *et cetera*). Assume that the population of forecasters has K different types, and recall that $h(\mathbf{z}_i|\boldsymbol{\theta}_i)$ subsumes differences in the parameterization of both the $h(\cdot)$ and the $F(\cdot)$ functions. At the first stage of the hierarchy of distributions, the i-th forecaster's response is characterized by $h(\mathbf{z}_i|\boldsymbol{\theta}_i)$, where the parameters $\boldsymbol{\theta}_i$ are distributed according to

$$\boldsymbol{\theta}_i \overset{iid}{\sim} \begin{cases} \boldsymbol{\theta}_1^* & \text{with probability } p_1 \\ \vdots & \\ \boldsymbol{\theta}_K^* & \text{with probability } p_K \end{cases} \tag{4}$$

or, equivalently, $\boldsymbol{\theta}_i \sim G$ i.i.d. for

$$G(\boldsymbol{\theta}) = \sum_{k=1}^{K} p_k \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_k^*) \tag{5}$$

[5] The mixture is parametrized as $F(y|\boldsymbol{\theta}) = (1-\omega)\Phi(y|\mu, \sigma_1^2) + \omega\Phi(y|\mu + \mu_\delta, \sigma_2^2)$, where $\Phi(y|\mu, \sigma^2)$ denotes the normal CDF with location μ and variance σ^2. The Zadora distribution is parameterized as

$$h(\mathbf{z}|\boldsymbol{\theta}) \propto \left(\prod_{j=1}^{J} \alpha_j(\boldsymbol{\theta})^{\xi_j} (1 - \alpha_j(\boldsymbol{\theta}))^{1-\xi_j} \right) \frac{\Gamma\left(\sum_{j \in \mathcal{J}^*} \phi\nu_j(\boldsymbol{\theta})\right)}{\prod_{j \in \mathcal{J}^*} \Gamma(\phi\nu_j(\boldsymbol{\theta}))} \prod_{j \in \mathcal{J}^*} z_j^{\phi\nu_j(\boldsymbol{\theta})-1},$$

where the auxiliary variable ξ_j is equal to 1 if $z_j = 0$ and 0 otherwise, $\mathcal{J}^* = \{j = 1, \ldots, J; \xi_j = 0\}$ indicates the bins with non-zero probabilities, $(\alpha_1(\boldsymbol{\theta}), \ldots, \alpha_J(\boldsymbol{\theta}))$ are the probabilities that a forecaster will report a zero on the J bins, and

$$\alpha_j(\boldsymbol{\theta}) = \int_0^\epsilon g(x|\nu_j(\boldsymbol{\theta}), r)dx,$$

where $g(x|m, r)$ denotes the PDF of a beta distribution $\mathcal{B}e(m, 100)$ with mean m and precision 100.

with $\delta(x)$ a point mass distribution located at 0, $p_k > 0$ and $p_1 + \ldots + p_K = 1$. At the second stage of the hierarchy, it is assumed that the unknown parameter types are sampled from a common distribution $\boldsymbol{\theta}_k^* \overset{iid}{\sim} G_0(\boldsymbol{\theta})$, $k = 1, \ldots, K$, and the type probabilities have prior distribution

$$(p_1, \ldots, p_K) \sim \text{Dir}\Big(\frac{\psi_0}{K}, \ldots, \frac{\psi_0}{K}\Big), \tag{6}$$

where ψ_0 is a concentration parameter and $\text{Dir}(a_1, \ldots, a_K)$ a Dirichlet distribution of parameters $(a_1, \ldots, a_K)$.

When K goes to infinity, one obtains

$$G(\boldsymbol{\theta}) = \sum_{k=1}^{\infty} p_k \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_k^*), \tag{7}$$

where $\boldsymbol{\theta}_k^* \overset{iid}{\sim} G_0$ and p_ks is a sequence of random weights with stick breaking representation $SB(\psi_0)$ described, for example, in Pitman (2006). This hierarchical model is known as a Dirichlet process prior:

$$\boldsymbol{\theta}_i \overset{iid}{\sim} G, \quad G \sim \mathcal{DP}(\psi_0, G_0),$$

where the random probability measure G is a Dirichlet process with law $\mathcal{DP}(\psi, G_0)$ (see Ferguson, 1973). The base measure G_0 has the interpretation of mean type distribution, and the precision parameter ψ_0 measures the concentration of G around G_0, so that when $\psi_0 \to +\infty$ all forecasters are assumed to be of the same type and when $\psi_0 \to 0$ the inference is done forecaster by forecaster (using the same prior).

The intuition behind Bayesian non-parametrics is that each forecaster can be described by a prior distribution over a sufficiently rich parameter space. Bayesian non-parametrics allows for some degree of pooling: the approach allocates forecasters whose predictive distributions are similar to one another into groups and allows the number of groups to grow naturally as more data becomes available. This pooling mitigates overfitting and produces sharper inference. At the same time, the non-parametric nature of the prior overcomes the inherent misspecification implied by the use of a specific parametric distribution and thereby delivers some consistency results when n goes to infinity which we briefly mention below.[6]

[6]The Dirichlet process prior is an example of a Bayesian non-parametric model and it has been applied in many fields, including econometrics (e.g., see Hirano, 2002; Griffin and Steel, 2011; Bassetti et al., 2014, 2018; Griffin and Kalli, 2018; Billio et al., 2019) and psychometrics (e.g., see Navarro et al., 2006; Karabatsos and Walker, 2009; Li et al., 2019). In econometrics, Bayesian non-parametrics has been used to build robust models that account for heavy tails, skewness, and multimodality. In psychometrics, Bayesian non-parametrics has been successfully employed to accommodate heterogeneity in experimental responses within models for cognitive processes.

Some asymptotic properties

Asymptotic properties of the proposed Bayesian model can be studied in two settings: large n and large J. The first asymptotic scenario concerns the consistency of the posterior distribution of the Bayesian model for the inference on the z's, while the second concerns the inference on the underlying probability distribution $F(\cdot)$. When the number of forecasters n grows to infinity, the posterior distribution concentrates around the process that is generating the data $\mathbf{z}_1, \mathbf{z}_2, \ldots$ (Ghosh and Ramamoorthi, 2003). Very loosely speaking, this implies that whatever the data generating process for the $\mathbf{z}_1, \mathbf{z}_2, \ldots$, the posterior will recover it. When, in addition, the number of bins J goes to infinity, the bin size goes to zero and the rounding disappears. Thereafter, the random histogram model $\mathbf{z}_i$ converges to an infinite dimensional model where each forecaster's response is modeled by a (random) CDF $Z_{i,\infty}$ with mean $F(\cdot|\boldsymbol{\theta}_i)$. Del Negro et al. (2018) provide a much more formal treatment.

Finite sample properties and caveats

Given a finite sample of forecasters, the posterior distribution can be easily approximated by Monte Carlo sampling (see Del Negro et al., 2018, for further details). This method generates random draws from the posterior distribution of $\boldsymbol{\theta}_i$ and of the subjective CDF $F(y|\boldsymbol{\theta}_i)$. For each posterior draw, the quantity of interest is computed, whether it involves individual forecasters or the cross-sectional distribution of forecasters. Point estimates, whenever used, and posterior credible intervals for any such object are obtained from this simulated posterior distribution.

The choice of the probabilistic model and of the prior distribution can have an impact on the results given that n is far from infinity (it hovers around 30 for the SPF) and especially because the number of bins J and of observations for forecasters are small (e.g., $J = 10$ in the US SPF on GDP in 2020). Therefore the choice of the probability density $h(\cdot)$, and in particular the distribution family $F(\cdot)$, does matter. When the number of bins decreases and/or the bin width increases, the amount of information available to reconstruct the subjective CDF diminishes, and model assumptions can have a more significant impact on the empirical results. An advantage of the Bayesian approach is that it accounts for the lack of information returning wider credible intervals, as we will see below. In general, the approach provides a measure of the level of estimation uncertainty for all objects of interest. Nevertheless, a robustness check with respect to the specification of the prior distribution and the distribution family should be considered in all applications of this method.

A comparison with existing approaches

We conclude this section with some comparisons between the different approaches in our application.[7] Figure 3 shows the inference results for the four SPF respondents shown in Figure 1. For each forecaster we show posterior draws (thin gray lines) from the BNP model for the subjective CDFs (top) and their quantiles (bottom), and compare it with the beta (black, dash-and-dotted lines for the CDF, and squares for the quintiles) and Gaussian (gray, dashed lines for the CDF, and circles for the quintiles) approaches. The CDF plots also show the observed cumulated histogram Z_{ij} (crosses), as in Figure 2. The quantiles shown in the bottom panels are the 5th, 10th, 25th, 50th, 75th, 90th, and 95th, with colors becoming lighter the farther away the quantile is from the median.

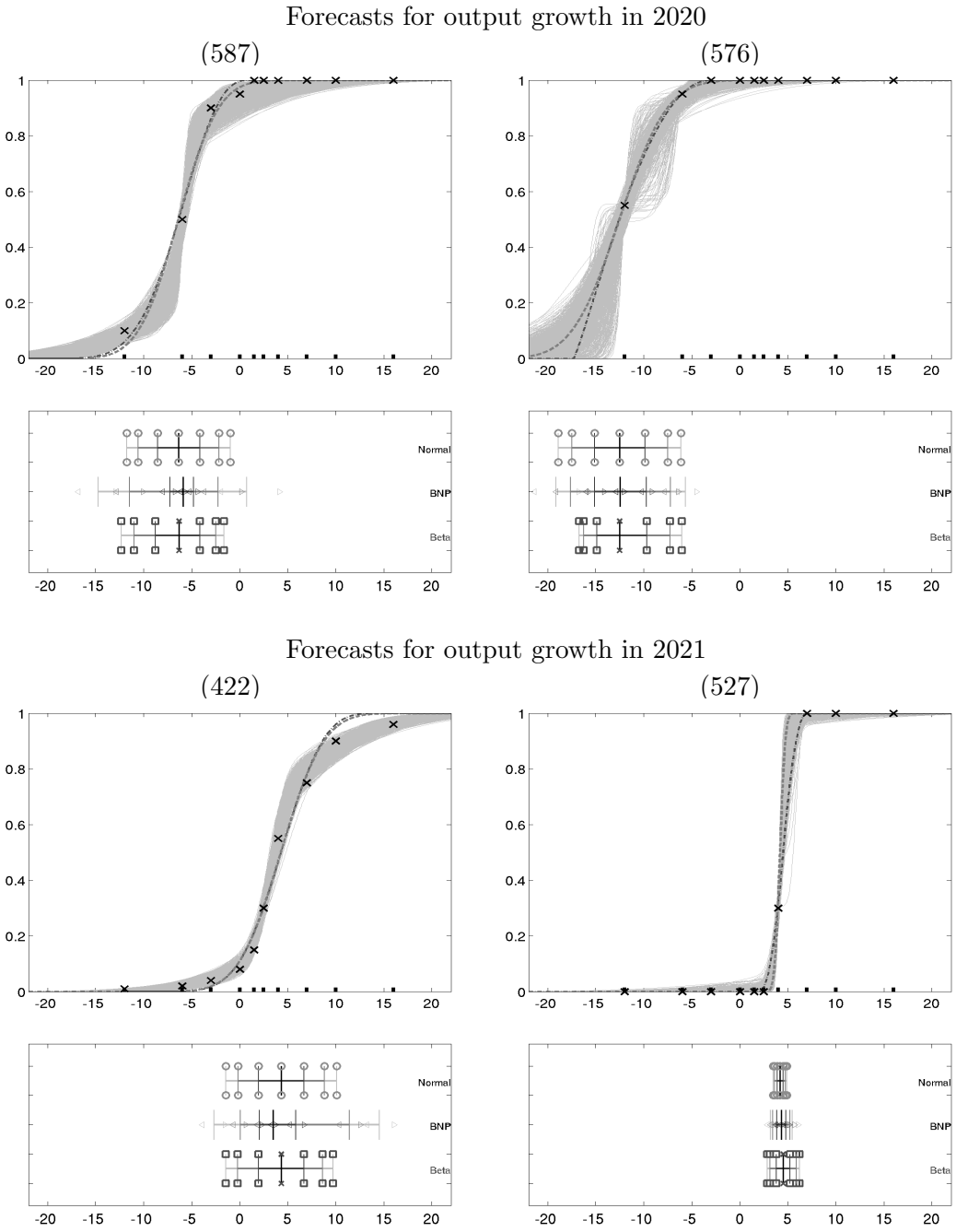
For the BNP approach, the bars represent the posterior means for each quantile and, for each quantile, the (5,95) posterior coverage intervals are shown using triangles. Figure 4 shows posterior draws from the BNP model for the subjective PDFs for the same forecasters, along with the step-wise uniform PDF obtained from histogram probabilities z_{ij}, $j = 1, \ldots, J$ (dotted lines).

Figure 3 is helpful in illustrating a few points about the BNP approach. First, the observed cumulative histogram (the Z_{ij}'s; crosses) belongs to the high posterior density region for all respondents. In contrast, as noted in discussing Figure 2, the beta and the normal approaches do not always fit the Z_{ij}'s well, and in these cases their CDFs do not belong to the high posterior density region obtained from the BNP approach. This implies that there can be significant differences in the quantiles implied by the different approaches (see, for instance, forecaster 587 or 422). Both the beta and the normal distributions miss the fat-tails of the distribution (e.g., see the left tails in the CDF chart, which are apparent also from Figure 4), and the interquartile range is much wider than that implied by the BNP approach.

Figure 3 also shows that whenever there is less information from the respondent, the BNP approach delivers wider posterior coverage intervals reflecting the higher degree of uncertainty. The case of respondent 576 is exemplary. This respondent places 55 percent

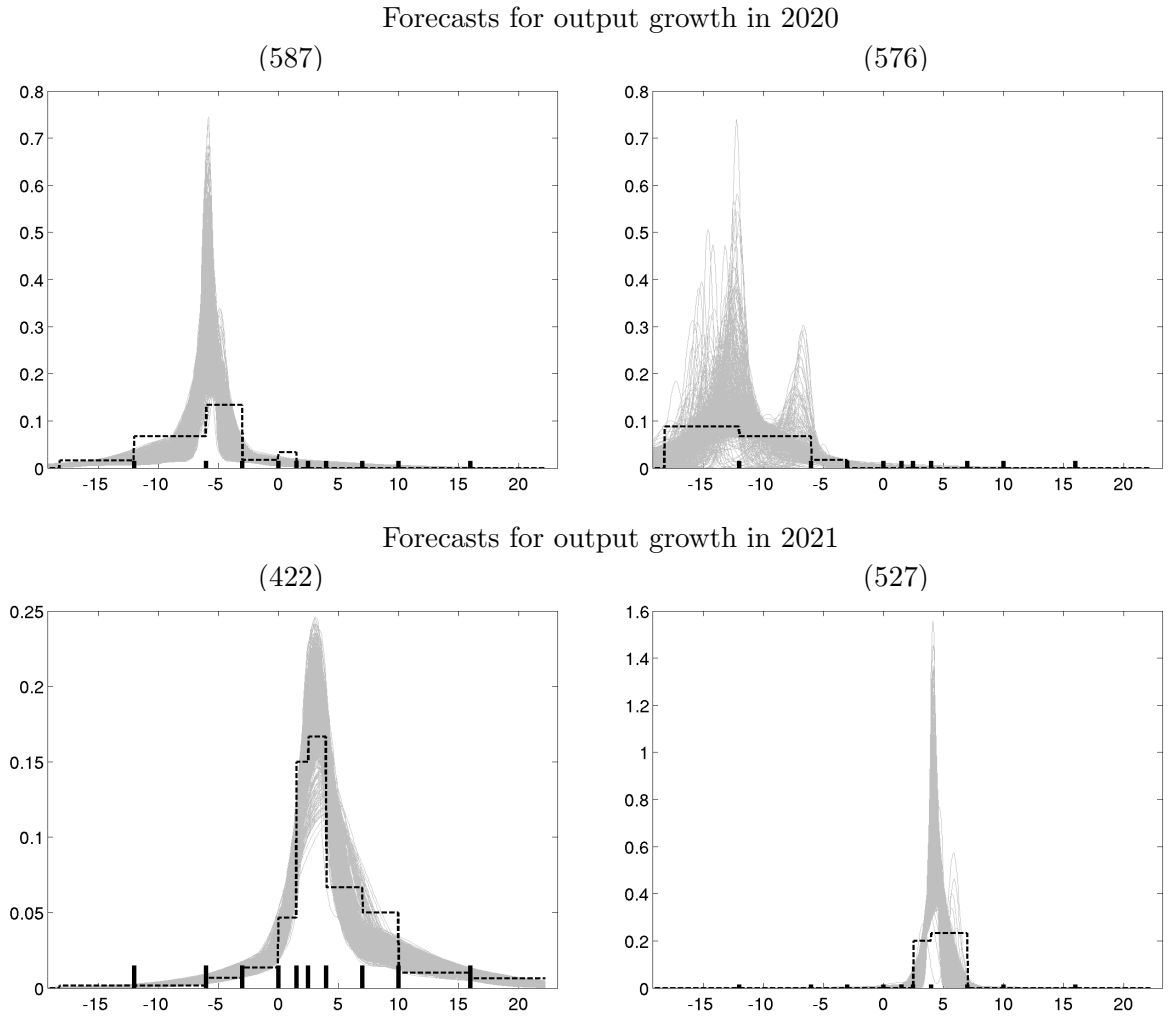
[7] The prior distribution on the parameters μ, μ_δ, σ_1, σ_2, ω, ϕ and ϵ is a Dirichlet process with concentration $\psi_0 = 1$ and base measure G_0 given by $\mu \sim \mathcal{N}(2, 5^2)$ (normal); $\sigma_j \sim \mathcal{IG}a(a_\sigma, b_\sigma)\mathcal{I}(\sigma_j)_{(0,10)}$ $j = 1, 2$ (inverse gamma), $E[\sigma_j] = 2$ and $V[\sigma_j] = 4$; $\mu_\delta \sim \mathcal{N}(0, 2.5^2)$ (normal); $\omega \sim \mathcal{B}e(0.5, 3)$ (beta); $\phi \sim \mathcal{G}a(a_\phi, b_\phi)\mathcal{I}(\sigma_j)_{(0,10)}$ (gamma), s.t. $E[\phi] = 25$ and $V[\phi] = 100$; $\epsilon \sim \mathcal{G}a(a_\epsilon, b_\epsilon)$ (gamma) s.t. α_j is close to one for $\nu_j < 0.01$, very small for any $\nu_j > 0.05$, and virtually zero when $\nu_j > 0.1$.

Figure 3: Inference Using Bayesian Non-parametric Approach: CDFs and Quantiles for Selected Examples from the 2020Q2 SPF



Note: In each panel: the subjective CDFs (top panels) and selected quantiles (bottom panels). Top panels: subjective CDF using least-squares approach with normal (gray, dashed line) or beta (black, dash-and-dotted line) assumption; subjective CDF using BNP approach (posterior random draws in light gray); and observed cumulated histogram probabilities Z_{ij} $j = 1, \ldots, J$ (crosses). Bottom panels: Quantiles of the predictive distribution computed using the normal (gray circles) or beta (black squares), and the BNP approach. The quantiles shown are the 5th, 10th, 25th, 50th, 75th, 90th, and 95th, with colors becoming lighter the farther away the quantile is from the median. For the BNP approach the bars represent the posterior means for each quantile and, for each quantile, the (5,95) posterior coverage intervals are shown using triangles.

Figure 4: Inference Using Bayesian Non-parametric Approach: PDFs for Selected Examples from the 2020Q2 SPF

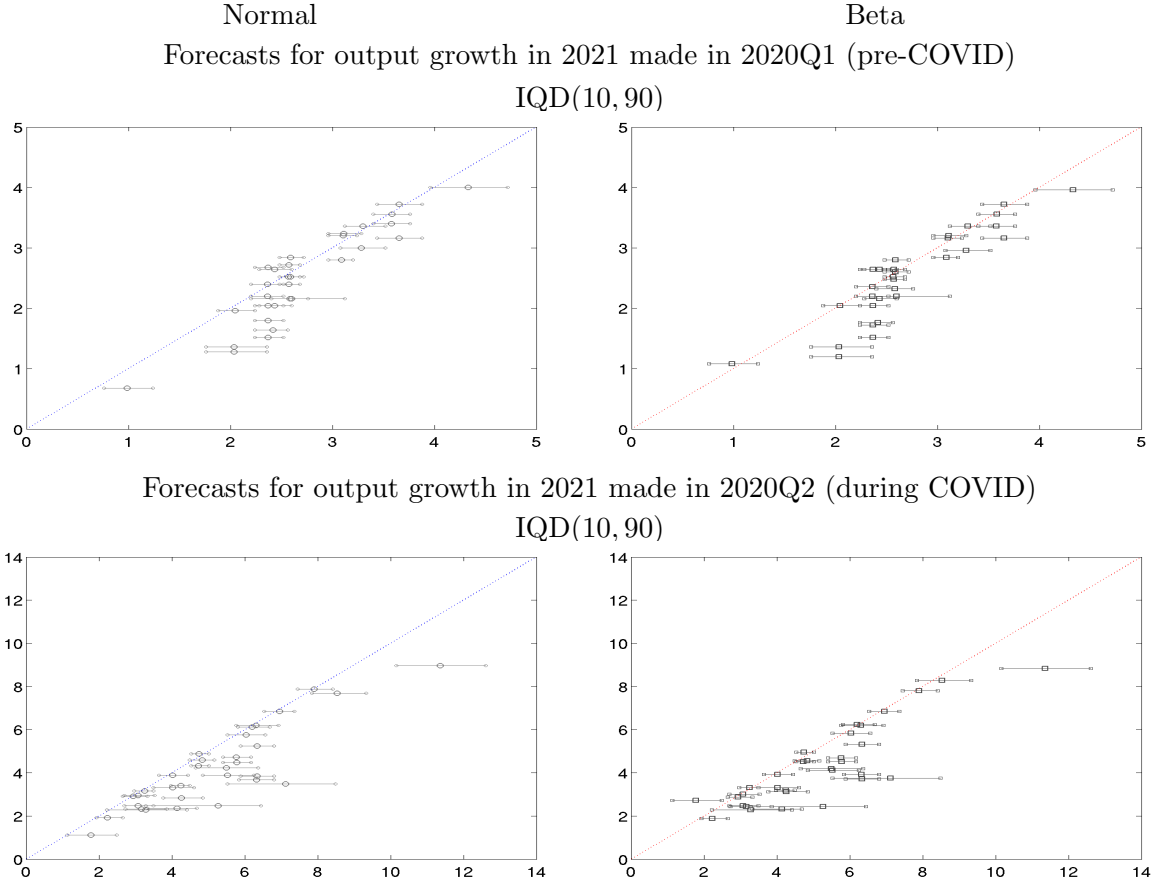


Note: In each panel: subjective PDF using BNP approach (posterior random draws in gray); and step-wise uniform PDF (dotted lines) implied by the histogram probabilities z_{ij}, $j = 1, \ldots, J$.

probability on the left open bin (see Figure 1), implying that we know very little about the left-tail behavior of this forecaster. The posterior coverage intervals for both the BNP CDF and PDF reflect this uncertainty, as evidenced by the fact that the gray lines for both the CDF and the PDF are far less concentrated for forecaster 576 in the left tail than for other forecasters (e.g., 587). Moreover, the gray lines for the left tails are less concentrated than for the right tails, where we have more information. The posterior distribution for the quantiles (especially for the left tail) also reflects this lack of information.

While Figures 2, 3, and 4 looked at a few examples, Figure 5 compares all participants across different parametric assumptions and estimation approaches, for the forecasts for 2021

Figure 5: Inter-quantile Ranges: Implications from Alternative Approaches



Note: In each plot, the dot coordinates represent the IQDs (or IQR) computed according to the BNP (horizontal axis) and the least-squares (vertical axis) approaches for the normal (left, gray circles) or beta (right, black squares) distributions.

made in 2020Q1 and 2020Q2.[8] In the comparison we use the interquantile distances (IQD) of the subjective CDF. Let F denote a CDF; its IQD(q, r) is defined as

$$IQD(q, r) = F^{-1}(r/100) - F^{-1}(q/100) \tag{8}$$

with $q \leq r$ and $q, r \in [0, 100]$. The interquartile range (IQR), a measure widely used to measure uncertainty in probability forecasts (see ClementsetalCHAPT-inVOLUME, 2022), is a special case of the IQD(q, r) for $q = 25$ and $r = 75$.

Each plot shows the IQDs computed according to BNP on the x-axis versus the IQDs computed using the normal (left column; gray circles) or beta (right column; black squares) approaches on the y-axis. The dot reports the BNP posterior mean while the thin horizontal

[8]Note that beginning with the 2020:Q2 SPF, changes were made to the definition of the bins for real GDP growth. Prior to 2020Q2, there were 11 bins with bounds −3, −2, −1, 0, 1, 2, 3, 4, 5, and 6 with left and right open bins.

lines provide the 95 percent posterior credible intervals. The 45-degree line indicates BNP and least-squares estimates are equal. While some of the points lie not too far from the 45-degree line, there are several deviations. For many forecasters, the BNP estimates of the IQD$(10, 90)$ are smaller than the least-square estimates (dots below the 45-degree line), suggesting that the subjective CDF BNP estimates' tails have a slower decay rate than one of the other models. Note that this is the case both before and after COVID, although the range of the x-axis is very different. For the IQRs (not shown), differences between BNP and alternative approaches are smaller—most points lie not far from the 45-degree line. Still, if we compute the ratio of the IQRs computed using the different approaches, we find that the posterior mean of these ratios varies from below 0.5 to above 1.5, suggesting that different approaches can lead to a different assessment of uncertainty.

III Challenges in Measuring Uncertainty

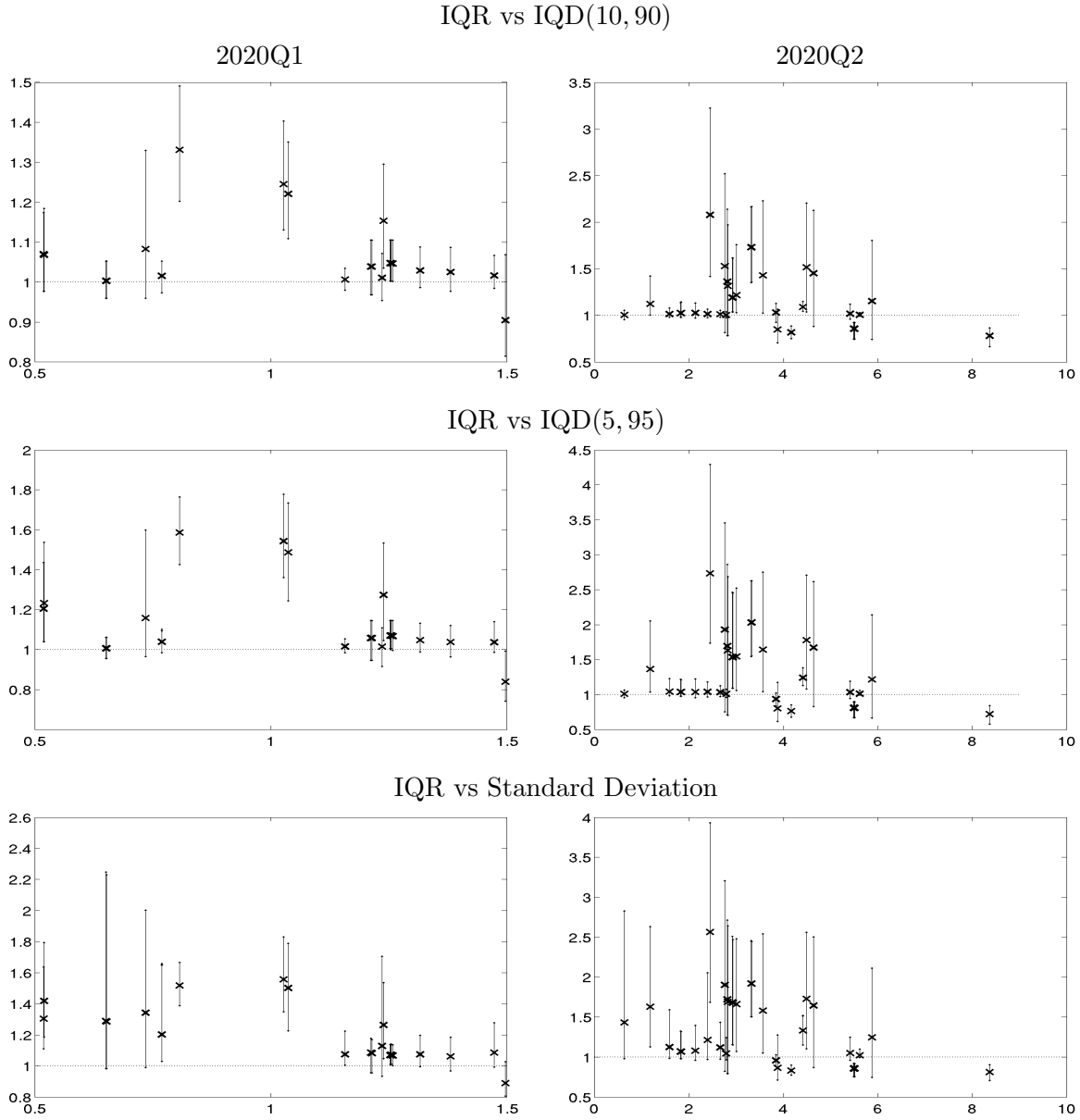
Assessing uncertainty plays an important role in macroeconomics analysis (Bloom, 2009, and Jurado et al., 2015, are two notable examples, and Bloom, 2014, provides a survey). Density predictions have the advantage over point predictions in that they provide information about the forecasters' subjective view on uncertainty. A large body of literature, much of it reviewed in ClementsetalCHAPT-inVOLUME (2022), makes use of uncertainty measures based on surveys of forecasters and investigates their properties.[9]

Measuring uncertainty in the context of density forecasts is not a trivial task, and this section reviews the approaches that have been used in the literature and the extent to which they may lead to different answers (see also Manski, 2018, for an insightful discussion). The variance (or the standard deviation) is a natural measure of uncertainty, but it is sensitive to assumptions on the tail behavior of the PDF. Cognizant of its current limitations (e.g., the fact that the beta chops off the tails altogether, or that the Gaussian does not capture possible fat tails), the literature has often relied on measures of uncertainty based on the IQR or other IQDs (e.g., Engelberg et al., 2011; Manski, 2018; Bruine De Bruin et al., 2011, among many others). If one ignores issues of rounding, IQRs are particularly robust because

[9] Zarnowitz and Lambros (1987) and Giordani and Soderlind (2003) find that forecasters underestimate uncertainty. Boero et al. (2008a),Daniel and Hirshleifer (2015), Kenny et al. (2014), Malmendier and Taylor (2015) also find that forecasters are overconfident and discuss various strategy to measure overconfidence. Some researchers (e.g., Liu and Sheng, 2019) explore the impact of subjective uncertainty on macroeconomic activity or discuss its evolution over time (e.g., Campbell, 2007).

the bins' edges can provide a non-parametric boundary for the quartiles. Since in normal circumstances it is quite rare for respondents to place more than 25 percent probability on the open bins, if $F(y_j) < .25$ and $F(y_{j+1}) > .25$, then it had better be that $y_j < F^{-1}(.25) < y_{j+1}$.

Figure 6: Which Measure of Uncertainty?



Note: In each plot: dots coordinates are the ratio between the BNP IQR/IQD ratio and the theoretical value of the ratio under Gaussianity (vertical axis) and the mean IQR (horizontal axis); vertical lines represent credible intervals.

A number of challenges remain, however. First, rounding may make non-parametric boundaries less reliable. Second, it is not clear what to do in the event that the given parametric distribution does not fit the points in the CDF well, such that $F^{-1}(.25)$ or $F^{-1}(.75)$

and the non-parametric boundaries are at odds— the beta for forecaster 422 in Figure 2 being one example. Third, there are cases, as we have seen, where respondents place more than 25 percent probability on the outer bins, and these cases are more frequent precisely when the uncertainty increases, that is, when monitoring it becomes most important. Fourth, when the gap between bins is wide, the IQR remains dependent on the assumed parametric distribution, as is the case for output forecasts during COVID. For instance, $F^{-1}(.25)$ could be anywhere between -6 and -12 percent for respondent 587 in Figure 2.

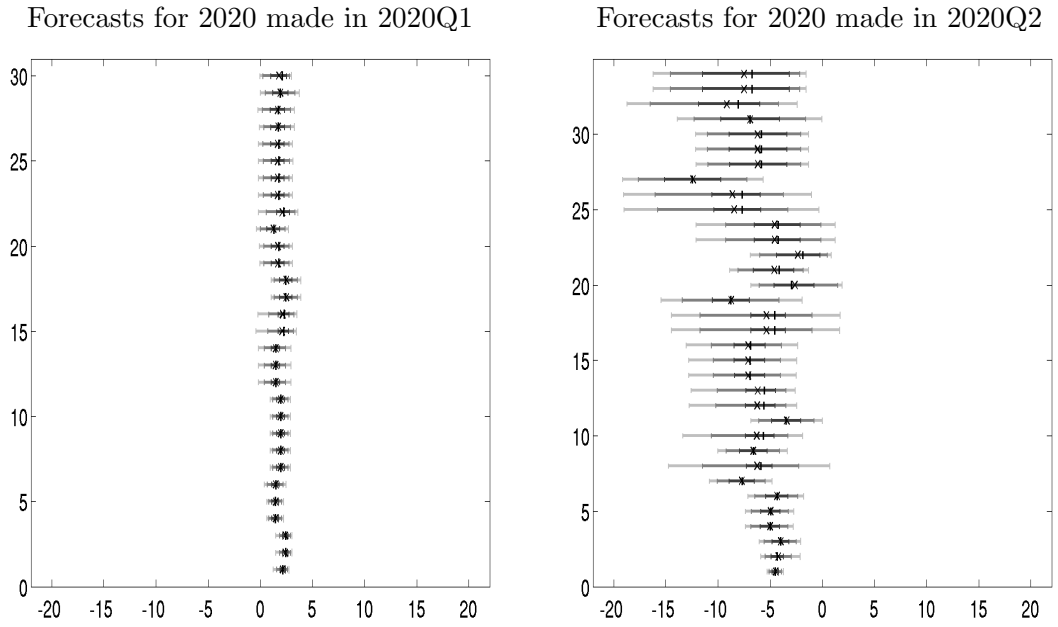
Perhaps the most important issue is that only under Gaussianity are the IQR, other IQDs, and the standard deviation proportional to one another and provide the same assessment of uncertainty. In the presence of skewness or kurtosis these various measures can provide quite different answers. We have seen above examples of distributions that appear to be far from Gaussian. How general is this situation? In the remainder of the section, we use the SPF dataset to provide an illustration of this issue.

Figure 6 considers two sets of forecasts: forecasts for 2020 made before (2020Q1, left column) and after (2020Q2, right column) COVID. The first row plots show, for each respondent in the survey, the ratio (vertical axis) between the IQR and the IQD$(10, 90)$ (BNP estimates), divided by the theoretical value of this ratio under Gaussianity versus the mean IQR (horizontal axis). In other words, if the distributions were all Gaussian, all the points should be equal to 1 (horizontal dashed line). We see that under standard circumstances (left column) when uncertainty is small, most observations are fairly close to 1 with only a handful of exceptions. When uncertainty becomes very large, however (right column), deviations from Gaussianity become both larger (a ratio of 2 means that the IQD$(10, 90)$ is twice as large as that implied by the IQR under normality) and more common. The last two rows perform the same analysis for IQD$(5, 95)$ and for the standard deviation. Not surprisingly, the deviations from Gaussianity are even larger when looking at these two statistics and quantitatively striking for forecasts made in 2020Q2: for several forecasters, uncertainty is at least 50 percent larger when measured using the IQD$(5, 95)$ or the standard deviation relative to the IQR. The bottom line from this section is that one measure of uncertainty is probably not sufficient, particularly in periods when uncertainty is high. For this reason, when assessing the evolution of uncertainty over time in section VI we look at a variety of measures.

IV Heterogeneity in Density Forecasts

Subjective density forecasts can be very different from one another. The probabilistic predictions for output growth in 2021 made by forecasters 422 and 527 in Figure 1 are a shining example of this phenomenon. This section discusses heterogeneity in both the first and the second moments of the forecast distributions, using the SPF surveys during the pandemic as an illustration.

Figure 7: Quantiles and Mean of Individual Density Forecasts for Output Growth



Note: For each object (quantiles and mean) we report the posterior mean.

It was established long ago—at least since Wachtel (1977) and Bomberger and Frazer (1981)—that forecasters disagree about point forecasts, and a large literature has documented and tried to explain this phenomenon for macroeconomic surveys (see, for instance, Mankiw et al., 2003; Carroll, 2003; Capistrán and Timmermann, 2009; Patton and Timmermann, 2010, 2011; Andrade and Le Bihan, 2013; Andrade et al., 2016 and other papers surveyed in ClementsetalCHAPT-inVOLUME, 2022). While much of the early literature focused on point projections, some studies documented the fact that forecasters disagree about uncertainty (e.g., Lahiri and Liu, 2006, D'Amico and Orphanides, 2008 and Manski, 2018).[10] A number of papers provide evidence of persistent heterogeneity in subjective uncertainty (Bruine De Bruin et al., 2011; Boero et al., 2014) and interpret this fact as

[10] Lahiri and Liu (2006) measure the uncertainty by fitting a Gaussian CDF to each forecaster's histogram

suggesting that the degree of uncertainty is a forecaster-specific characteristic akin to the individual optimism and pessimism established in the literature on point forecasts. While the papers mentioned above provide quantitative evidence on forecasters' heterogeneity by focusing on the variance of first- or second-order moments, other work discusses higher moments (e.g. Mirkov and Steinhauer, 2018) or uses measures of discrepancy based on entire predictive distribution (e.g., Shoja and Soofi, 2017, Cumings-Menon et al., 2021, and Rich and Tracy, 2021). In particular, Rich and Tracy (2021) propose a measure of heterogeneity based on the Wasserstein distance, which they computed assuming that individual PDFs are step-wise Uniform distributions. See also ClementsetalCHAPT-inVOLUME (2022) for further references.

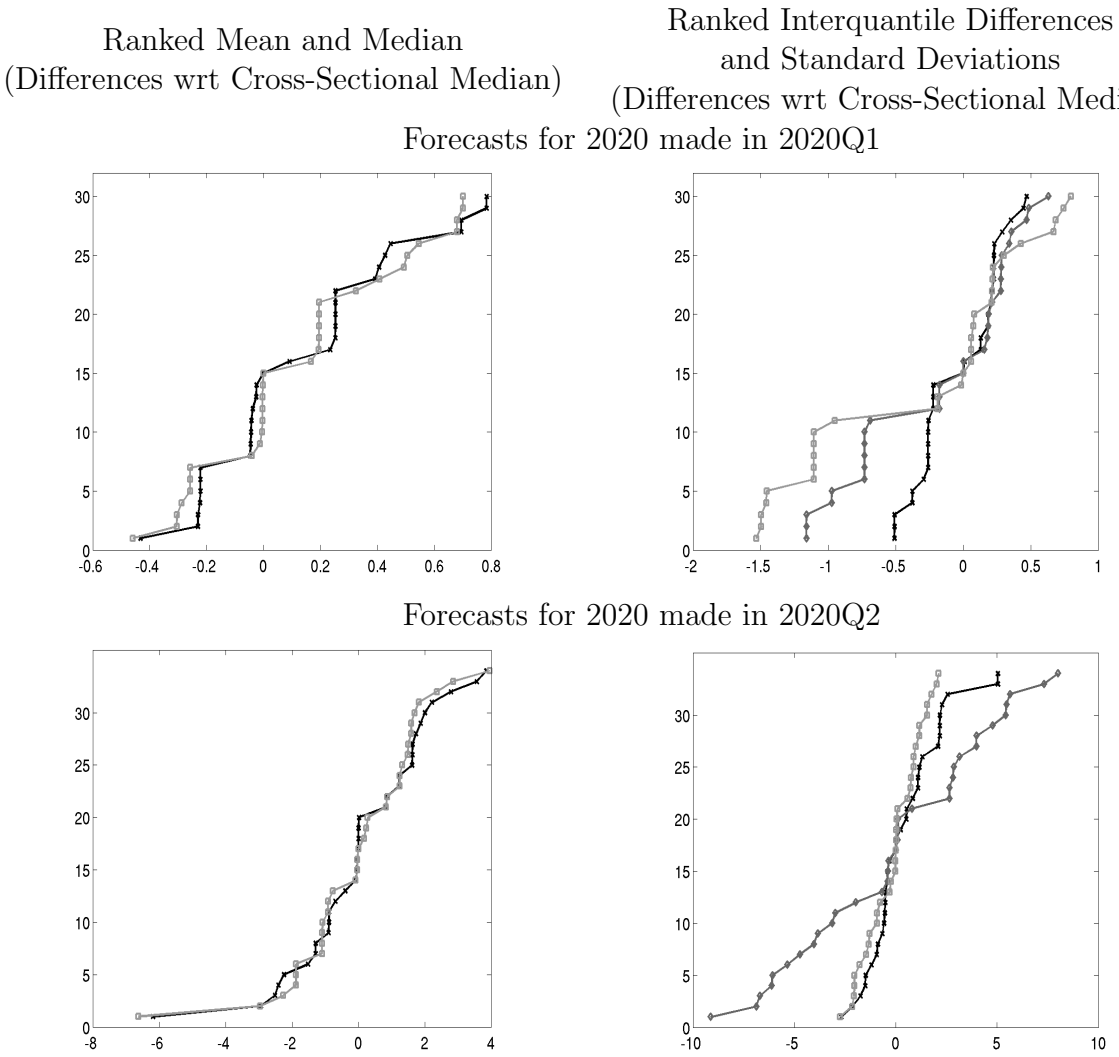
In the remainder of the section we document the fact that SPF forecasters do indeed significantly disagree about the uncertainty in the economy and that the extent of their disagreement increases in high-volatility periods such as COVID. Figure 7 shows the quantiles for the individual output growth density forecasts for 2020 made in 2020Q1 (left panel), right before the COVID pandemic reached the US, and the following quarter, 2020Q2 (right panel). For each forecaster the quantiles of the distribution are depicted using different shades of gray: light gray for the segment connecting the 5th and 95th quantiles, slightly darker gray for that connecting the 10th and 90th quantiles, and black for the 25th-75th range. The median is denoted using a single vertical black bar and the mean by a cross. For each object (quantiles and mean) we report only the posterior mean, because visualizing posterior uncertainty as in Figure 3 becomes challenging. In each panel the forecasters are sorted from 1 to n, where n is the total number of forecasters in each survey, on the basis of their (posterior mean) IQR.

Since the survey composition changes from survey to survey the figure does not allow for comparing the change in uncertainty for the same individual (some of the results shown later in Section VI are more suited for this purpose). The point of Figure 7 is instead twofold. First, it provides a bird's-eye view of the dramatic differences in uncertainty between the two surveys, even though their temporal distance is only one quarter (the x-axis is the same in

and using the "whisker plots" (that is, plotting some key quantiles of the distribution). D'Amico and Orphanides (2008) measure the individual variance under the Normal parametric assumption, assume a Gamma distribution for the cross-section of individual variances, and use the variance of this Gamma to measure disagreement about uncertainty. Still in the context of the SPF, Manski (2018) plots the median versus the IQR for a cross-section of SPF forecasters and thereby documents the heterogeneity in both the central tendencies and uncertainty. a few papers (Clements, 2014b; Manzan, 2021, among others) discuss the updating of density forecasts and in particular uncertainty in light of new information.

both panels to emphasize this point). In 2020Q1 essentially all IQD(5,95) are between 0 and 4, while in 2020Q2, for many of the participants, they are several times as large. Second, it shows that heterogeneity in both central tendencies and subjective uncertainty is sizeable. Some forecasters remain quite confident about their predictions even during COVID, while for others the IQD(5,95) is almost as large as 20 percent. While the fact that the location of the bins changed from 2020Q1 to Q2 may explain part of the increase in average uncertainty across surveys, it is less clear that it has an effect on the increase in heterogeneity.

Figure 8: Disagreement About Mean, Median, and Uncertainty for Output Growth



Note: Left: ranked differences in the mean (light gray, squares) and the median (black, crosses). Right: ranked differences in the IQR (black, crosses), IQD(5,95) (gray, diamonds), and standard deviation (light gray, squares). The reported values are the posterior means of the quantities of interest.

Figure 8 provides evidence on the heterogeneity in central tendencies and uncertainty, as well as its evolution with COVID. The left column in Figure 8 shows the means (light gray,

squares) for all respondents, sorted in increasing order, and expressed in difference from the cross-sectional median (of the means). The black line with crosses does the same for the median. The right column uses the same approach to quantify heterogeneity in uncertainty, as measured by the IQR (black, crosses), the IQD$(95, 5)$ (gray, diamonds), and the standard deviation (light gray, squares). Full homogeneity implies that each line is vertical at zero. The flatter the curve, the higher the level of disagreement among forecasters. The top and bottom row concern predictions made in 2020Q1 and 2020Q2, respectively.

That the scale of the x-axis is almost one order of magnitude larger for 2020Q2 than 2020Q1 for both heterogeneity in central tendencies and uncertainty is quite telling. Differences in means or medians between more and less optimistic forecasters are on the order of about 1 percent in 2020Q1. They become about 10 percent in 2020Q2. Heterogeneity in uncertainty is more sensitive to how it is measured. In 2020Q1 the difference in IQRs between the least and the most uncertain respondent is on the order of 1 percent, while the difference in IQD(5,95) and standard deviations is about 1.5 and 2 percent, respectively. In 2020Q2, as the COVID pandemic hit the US, the difference in IQRs between the least and the most uncertain respondent is about 7 percent, while that in IQD(5,95) is almost a staggering 20 percent. This suggests that the assessment of tail events is far more different in 2020Q2 than before COVID. The different behavior of the IQRs and IQD(5,95) confirms the departure from Gaussianity discussed in the previous section. The difference in standard deviations in 2020Q2 is smaller, however, and closer to that of IQRs.

An interesting feature of Figure 8 is that the clustering of the CDF forecasters in terms of both means and uncertainty is quite evident from the staircase-like behavior of the curves for the pre-COVID density predictions, with three or more forecasters having very similar central tendencies or degrees of uncertainty. In 2020Q2 the curves are more continuous, although there is still some amount of clustering. This also indicates that the amount of heterogeneity increases together with uncertainty during COVID.

V Pooling and Consensus Forecasts

It is common practice to aggregate forecasters, if only to simplify the presentation of the analysis and avoid reporting all the individual responses. The goal of forecast combination is to reduce the information of a pool of forecasts to a single combined forecast (e.g., see Timmermann, 2006). Combination may occur at two different levels: point forecasts or

probability forecasts. For example, the Philadelphia Fed reports the SPF data aggregated across forecasters in two ways: the median of the point forecasts and the average probability mass in each bin for the probability forecasts. In the remainder of the section we focus on combining probabilistic forecasts.

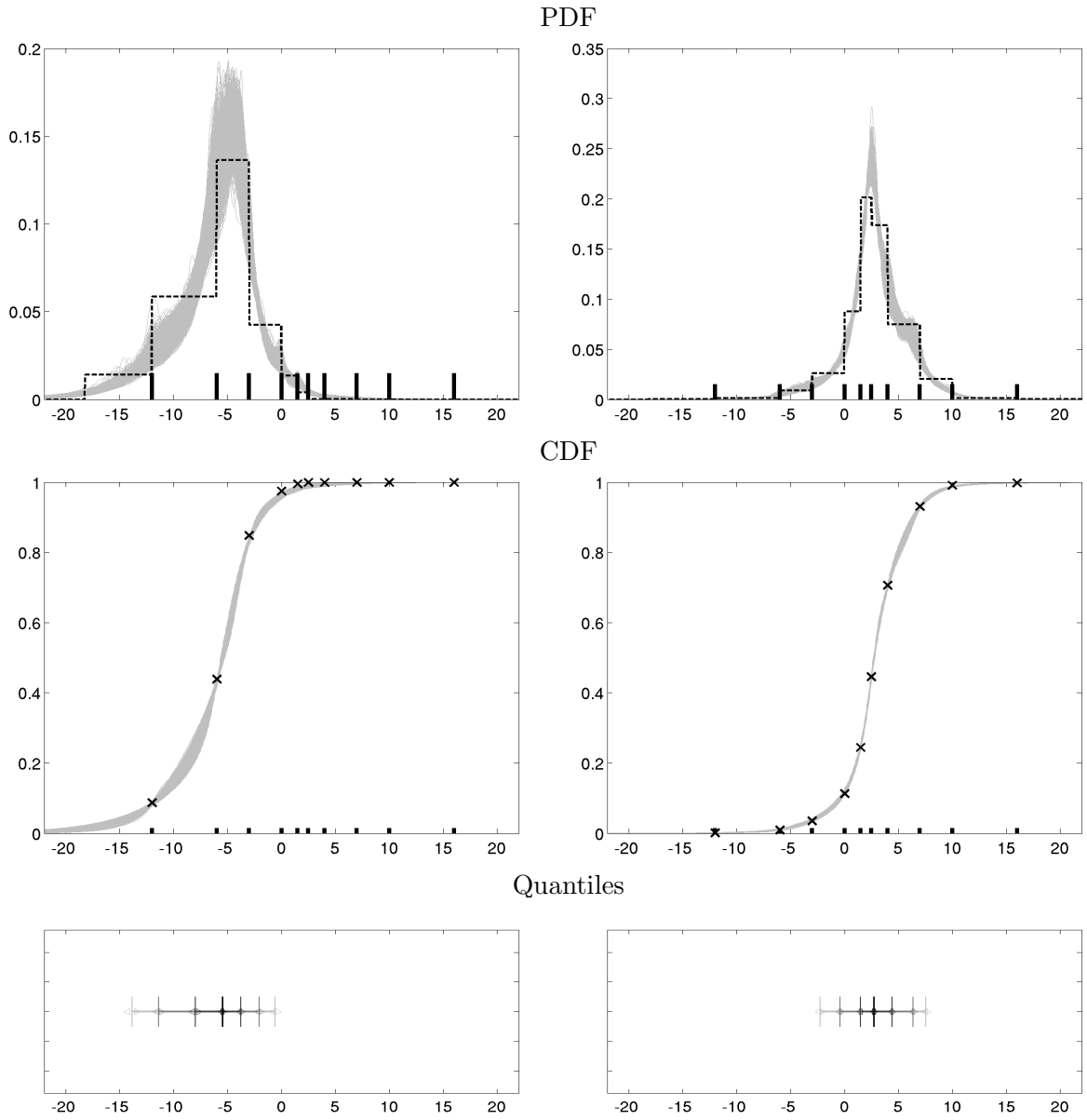
The simplest possible combination formula for histograms, densities, and CDF is the linear combination, also known as linear pooling (se, e.g. Genest and Zidek, 1986). The linear pooling of subjective PDFs and CDFs is defined as

$$\bar{f}^w(y) = \sum_{i=1}^{n} w_i f(y|\boldsymbol{\theta}_i), \quad \bar{F}^w(y) = \sum_{i=1}^{n} w_i F(y|\boldsymbol{\theta}_i), \tag{9}$$

where $w_i \geq 0$, $i = 1, \ldots, n$, $\sum_{i=1}^{n} w_i = 1$, and $f(y|\boldsymbol{\theta}_i)$ is the PDF associated with the subjective CDF $F(y|\boldsymbol{\theta}_i)$. The equally weighted linear pooling, also known as "consensus density forecast," is obtained for $w_i = 1/n$. In what follows, we denote with $\bar{f}(y)$ and $\bar{F}(y)$ the consensus PDF and CDF, respectively. Analogous definitions can be given for equally weighted linear pooling of histograms and cumulated histograms, that is, $\bar{z}_j = (z_{1j} + \ldots + z_{nj})/n$ and $\bar{Z}_j = (Z_{1j} + \ldots + Z_{nj})/n$. Empirical studies have shown that consensus performs relatively well in practice and usually outperforms more sophisticated pooling schemes (e.g., Zarnowitz, 1967, and more recently Genre et al., 2013, and Conflitti et al., 2015).

Figure 9 provides examples of consensus density forecasts. The top-row panels show the step-wise uniform PDF implied by the pooling probabilities $\bar{z}_j$, $j = 1, \ldots, J$, for output growth in 2020 (left) and 2021 (right) from the 2020Q2 SPF survey (step-wise dotted lines). They also show the estimated consensus PDF obtained using the BNP approach (posterior draws in gray). The middle-row panels show the cumulated histogram pooling probabilities $\bar{Z}_j$ $j = 1, \ldots, J$ (crosses) and the consensus CDF (posterior draws in gray). The bottom panels display selected quantiles of the consensus distribution computed using the BNP approach, with the bars representing the posterior means for each quantile and the triangles representing the (5%,95%) posterior coverage intervals. Figure 9 shows that the level of estimation uncertainty for the consensus PDF or CDF is generally much smaller than that for individual forecasters (recall Figure 3). This reduction in the inference uncertainty is an effect of both averaging and the fact that pooling favors posterior concentration. Figure 9 also suggests that estimation uncertainty for the consensus distribution in 2020 (left) is larger than that for 2021 (right), partly because in 2020 respondents assign substantial mass to the left open bin and to the very wide bin $(-12, -6]$.

Figure 9: 2020Q2 Consensus PDF, CDF, and Quantiles



Note: Top row: step-wise uniform PDF (dotted lines) implied by the pooling probabilities $\bar{z}_j$, $j = 1, \ldots, J$ and consensus PDF using BNP (posterior random draws in light gray). Middle row: cumulated histogram pooling probabilities $\bar{Z}_j$ $j = 1, \ldots, J$ (crosses) and consensus CDF using BNP (posterior random draws in gray). Bottom row: quantile ranges of the consensus distribution. Bottom panels: Selected quantiles of the consensus distribution computed using the BNP approach. The quantiles shown are the 5th, 10th, 25th, 50th, 75th, 90th, and 95th, with colors becoming lighter the farther away the quantile is from the median. The bars represent the posterior means for each quantile and, for each quantile, the (5%,95%) posterior coverage intervals are shown using triangles.

After the seminal paper by Stone (1961), pooling of densities has been extensively applied in the empirical literature (see, e.g. DeGroot and Mortera, 1991; DeGroot et al., 1995) and extended along different directions including nonlinear pooling (e.g., see Genest and Zidek,

1986; DeGroot and Mortera, 1991; Clements and Harvey, 2011), calibrated linear pooling (e.g., Ranjan and Gneiting, 2010), and generalized pooling (see, e.g. Kapetanios et al., 2015; Bassetti et al., 2018). In this literature, a parametric family of aggregation functions $C_{\boldsymbol{\eta}}$ from $[0,1]^n$ to $[0,1]$, with $\boldsymbol{\eta} \in E$, is used to pool together the CDFs:

$$\bar{F}^{\eta}(y) = C_{\boldsymbol{\eta}}(F(y|\boldsymbol{\theta}_1), \ldots, F(y|\boldsymbol{\theta}_n)). \tag{10}$$

The linear pooling is a special case of these general frameworks, which can be obtained choosing $C_{\boldsymbol{\eta}}(x_1, \ldots, x_n) = \eta_1 x_1 + \ldots + \eta_n x_n$ and $\eta_j = w_j$. Optimal pooling has been introduced to take advantage of the heterogeneous forecast abilities of the respondents (see, e.g. Hall and Mitchell, 2007; Geweke and Amisano, 2011; Clements and Harvey, 2011). In this framework the weights w_j, $j = 1, \ldots, n$, of the linear pooling or the parameters $\boldsymbol{\eta}$ of the generalized pooling are either specified as a function of the relative performances of the forecasters, or chosen following some statistical criteria. In both frameworks forecast performances are taken into account and usually obtained from previous forecasting exercises. The literature has also recognized that relative forecasting performance may change over time and has proposed time-varying pooling schemes (see, e.g. Billio et al., 2013; Del Negro et al., 2016; McAlinn and West, 2019).

The pooling techniques reviewed above make it possible to partly summarize the information content of the individual predictive densities. In the case of linear pooling it is possible to establish a relationship between the first and second moments of the pooled distribution and the cross-sectional distribution of the moments of the individual CDFs. For instance, the uncertainty in the linear pooling distribution can be related to the uncertainty in the individual forecasts. The variance of the consensus CDF, sometimes refereed to as *aggregate uncertainty*, is the sum of the *average (individual) uncertainty* $\bar{\sigma}^2$ and the *disagreement* among forecasters $\mathbb{V}(\mu)$, defined as

$$\bar{\sigma}^2 = \sum_{i=1}^{n} w_i \sigma_i^2, \quad \mathbb{V}(\mu) = \sum_{i=1}^{n} w_i (\mu_i - \bar{\mu})^2, \tag{11}$$

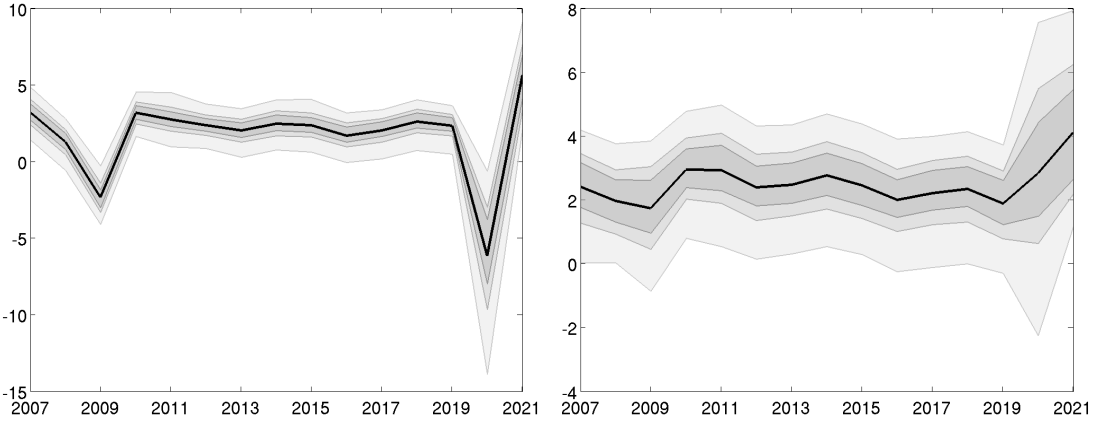
where μ_i and σ_i^2 are the mean and variance, respectively, of the i-th forecaster CDF, $i = 1, \ldots, n$, and $\bar{\mu} = w_1 \mu_1 + \ldots + w_n \mu_n$ is the mean of the linear pooling. These indicators have been used in macroeconomic analysis, to capture different aspects of the aggregate uncertainty (see, e.g. Giordani and Soderlind, 2003). See ClementsetalCHAPT-inVOLUME (2022) for more details on the uncertainty and disagreement decomposition and for a review of the literature.

VI The Evolution of Professional Forecasters' Density Forecasts during the COVID Pandemic

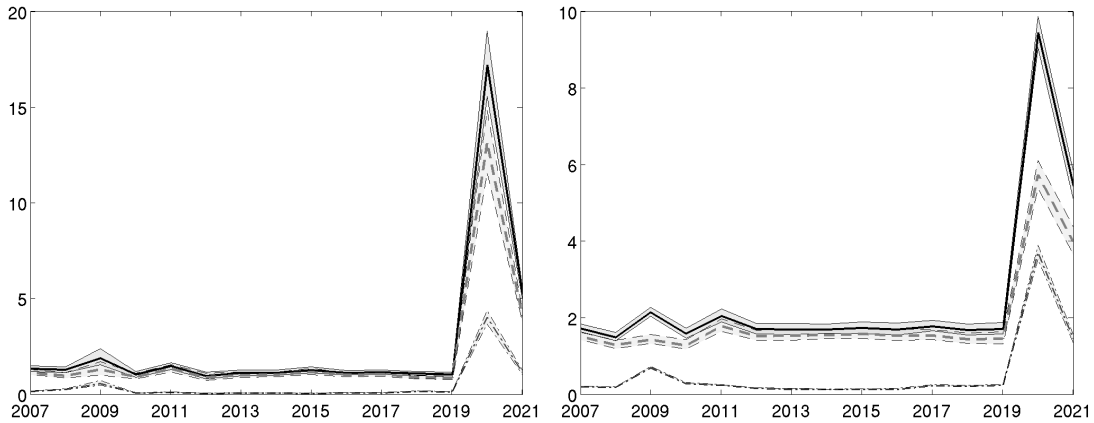
Figure 10: GDP Growth Uncertainty during COVID

Forecasts for Current Year | Forecasts for the Following Year

Mean and Quantiles of the Consensus Density Forecast



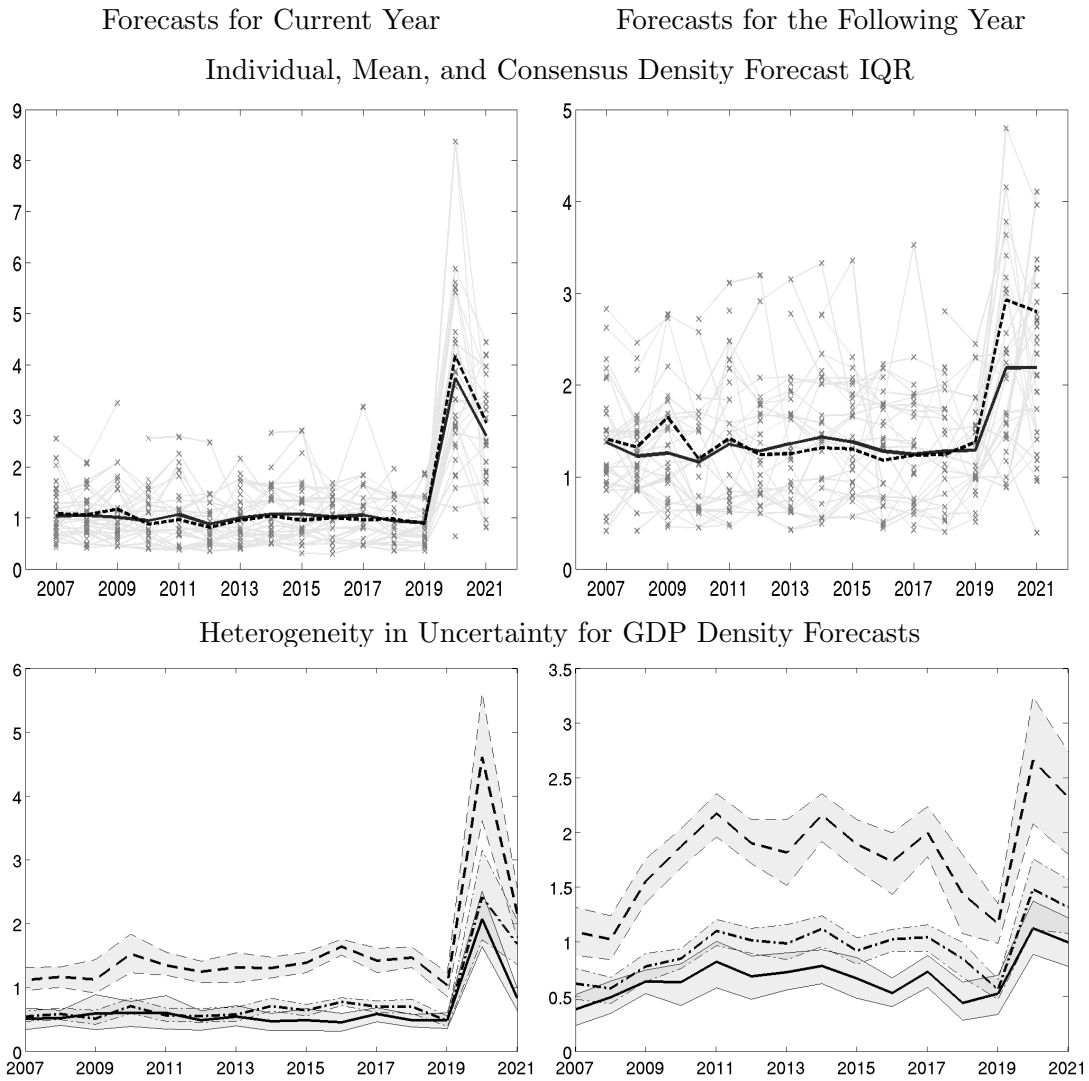
Variance of the Consensus Density Forecast, Average Uncertainty and Disagreement



Note: Left column: current year; right column: next year. Top row: mean of the consensus distribution (black solid line); the 5th, 10th, 25th, 75th, 90th, and 95th quantiles of the consensus distribution are displayed using different shades of gray. Bottom row: variance of the consensus forecast (solid black line), average uncertainty across forecasters $\bar{\sigma}^2$ (dashed gray line); disagreement $\mathbb{V}(\mu)$ (dash-and-dotted black line). For each object the shaded areas display the 90 percent posterior credible intervals.

This section describes the evolution of SPF density forecasts for GDP growth and inflation during the COVID pandemic. We discuss the change in the consensus forecast, in average measures of subjective uncertainty, as well as heterogeneity in both mean projections and subjective confidence in such projections. We show the time series for these various objects of interest from 2007 to 2021, where the starting date is chosen so that the analysis includes the Great Recession for comparison. We will focus on the surveys for both the

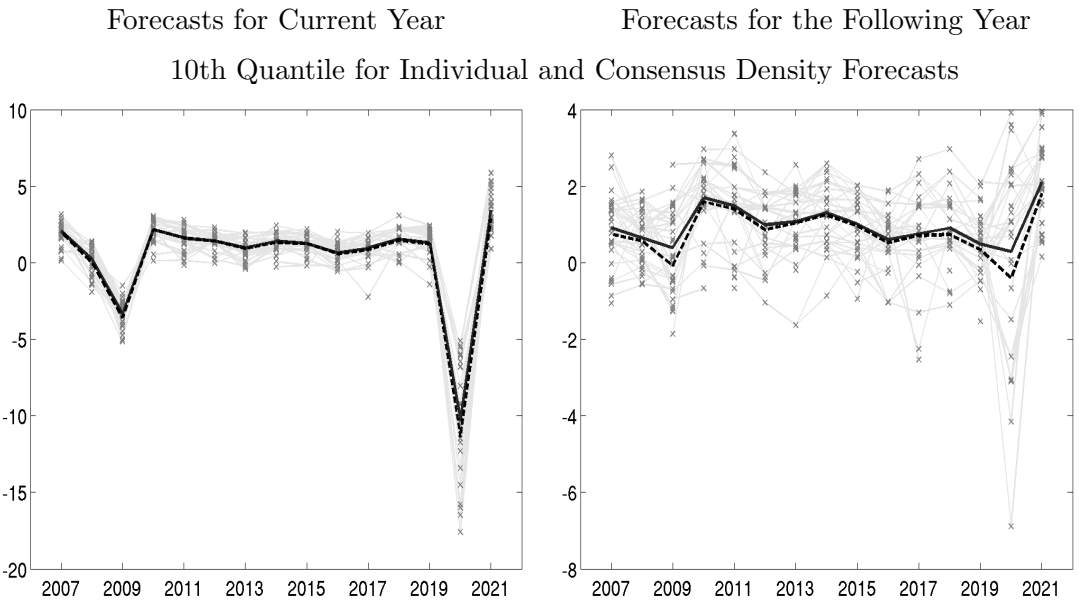
Figure 11: Average and Individual GDP Growth Uncertainty during COVID



Note: Top panels: Individual (light gray with crosses), mean (solid black), and consensus density forecast (dashed black) IQRs. Left column: current year; right column: next year. Bottom panels: Cross-sectional IQR for individual IQD(10,90)'s (dashed), IQRs (dash-and-dotted), and standard deviation (solid). For each object the shaded areas display the 90 percent posterior credible intervals. Left column: current year; right column: next year.

current and the following year made in Q2 of each year, keeping constant the quarter in which the survey is taken so that the forecast horizon is comparable across the time series. We choose Q2 because we have observations in 2020 after the pandemic had begun as well as for 2021, but the general conclusions are broadly similar regardless of the quarter chosen.

Figure 12: GDP Growth at Risk during COVID



Note: 10th quantile for individual (light gray with crosses) and consensus density forecast (dashed black); the solid black line depicts the average across individuals. Left column: current year; right column: next year.

VI.A GDP Growth

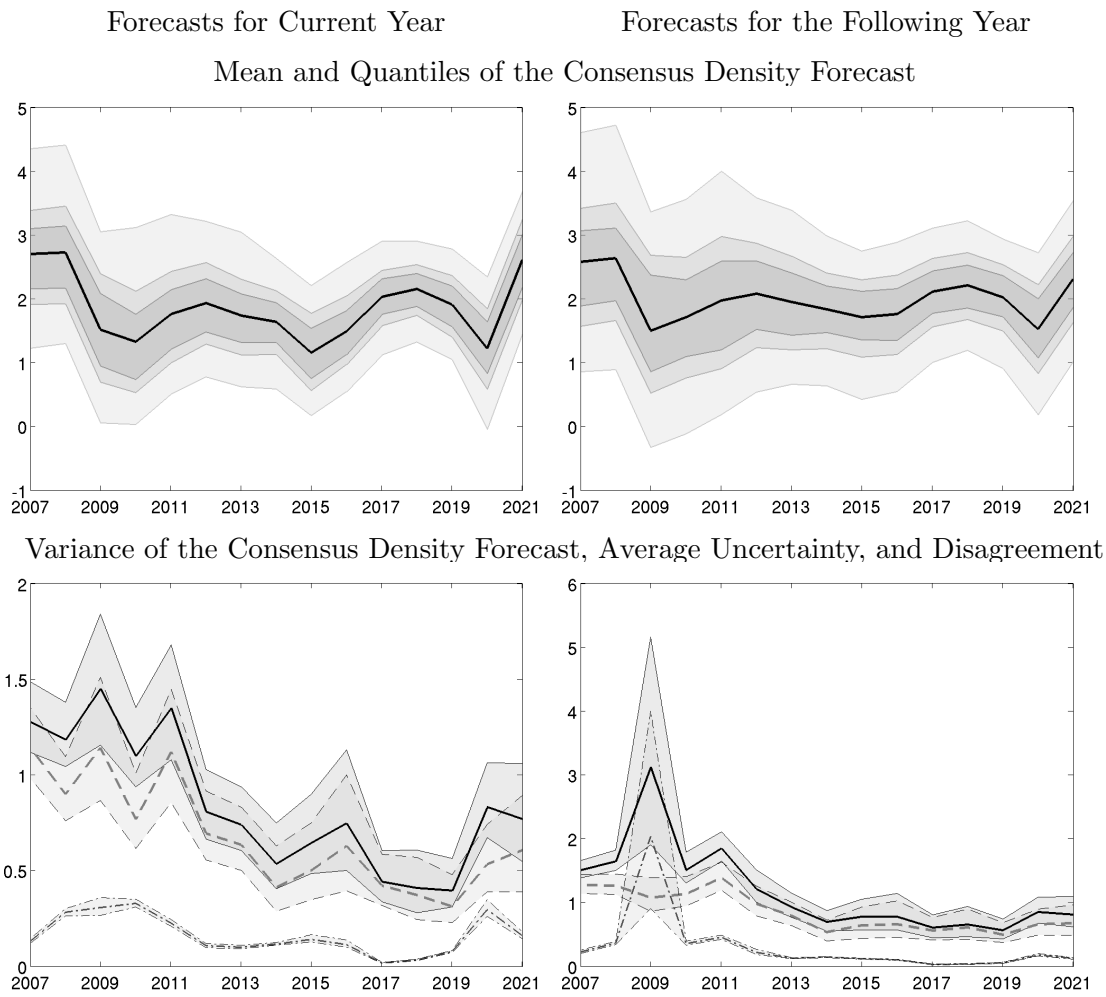
The two panels in the top row of Figure 10 display the mean and selected quantiles of the consensus forecast PDF $\bar{f}(y)$ defined in equation (9) for the current and the following year. Specifically, in each panel the black solid line in each panel shows the mean of the consensus distribution, while the 5th, 10th, 25th, 75th, 90th, and 95th quantiles of the distribution are displayed using different shades of gray (for all objects in the top panels we show the BNP posterior means).

The panels on the bottom row show the variance of the consensus forecast (solid black line) together with its decomposition between average uncertainty across forecasters $\bar{\sigma}^2$ (dashed gray line) and disagreement $\mathbb{V}(\mu)$ (dash-and-dotted black line). For each object the line represents the posterior BNP mean, while the shaded areas display the 90 percent posterior credible intervals.

Focusing on the projections for the current year (left column), we see that the consensus forecast distribution goes deep into negative territory in 2020, and its variance increases to unprecedented levels.[11] While disagreement also rises notably in 2020, most of the increase

[11] It is important to recall that the survey design, and in particular the location of the bin edges y_j, changed in 2009 and then again, dramatically, in 2020. This may have impacted the reported subjective uncertainty.

Figure 13: Inflation Uncertainty during COVID

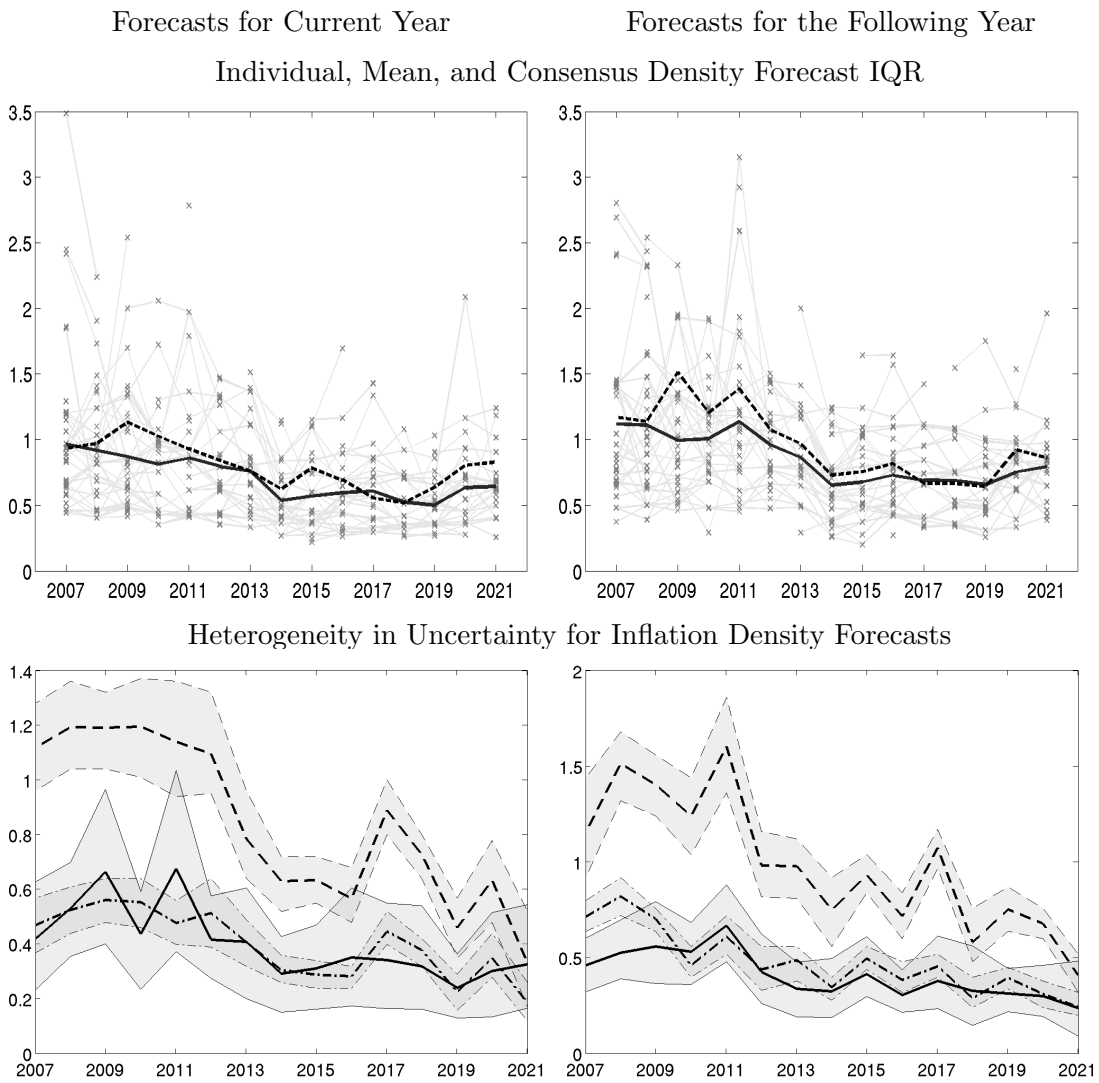


Note: Left column: current year; right column: next year. Top row: mean of the consensus distribution (black solid line); the 5th, 10th, 25th, 75th, 90th, and 95th quantiles of the consensus distribution are displayed using different shades of gray. Bottom row: variance of the consensus forecast (solid black line), average uncertainty across forecasters $\bar{\sigma}^2$ (dashed gray line), disagreement $\mathbb{V}(\mu)$ (dash-and-dotted black line). For each object the shaded areas display the 90 percent posterior credible intervals.

in the variance of the consensus distribution is due to the increase in the average variance across individuals. In 2021 the variance of the consensus forecasts, the average variance, and disagreement all fall to about one-third of their 2020 respective levels, but remain elevated compared to historical standards. The bottom panels also display the 90 percent credible intervals for each object of interest (shaded areas) and show that the increases in both uncertainty and disagreement during COVID are very significant even if one takes inference uncertainty into account.

Moving on to the forecasts for the following year (right column), the variance of the consensus forecast also rises to unprecedented levels in 2020, although it is smaller by about

Figure 14: Average and Individual Inflation Uncertainty during COVID

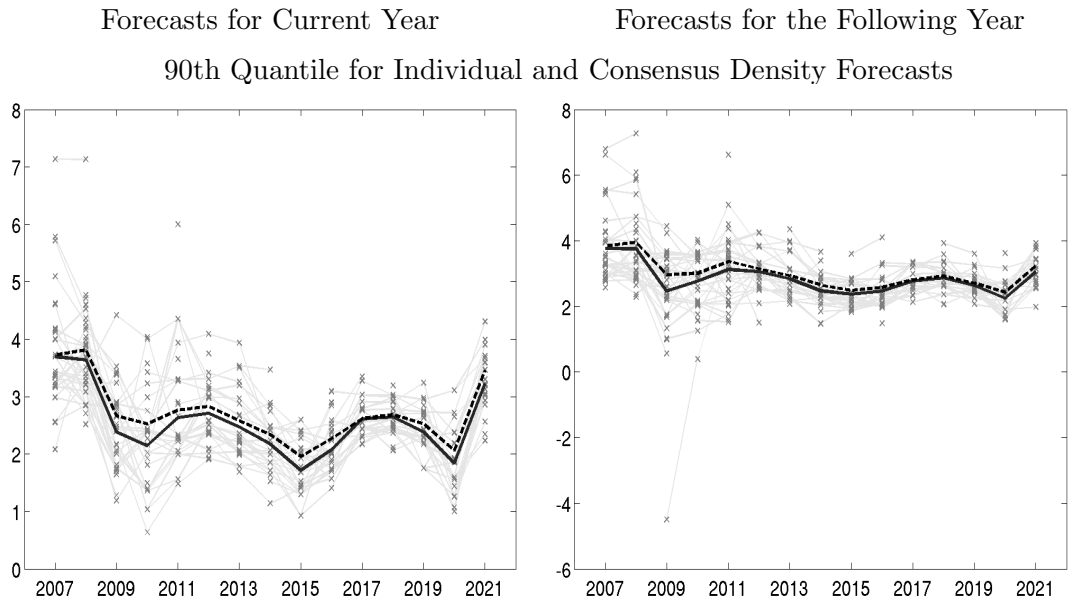


Note: Top panels: Individual (light gray with crosses), mean (solid black), and consensus density forecast (dashed black) IQRs (top row) and IQD(10,90) (bottom row). Left column: current year; right column: next year. Bottom panels: Cross-sectional IQR for individual IQD(10,90)'s (dashed), IQRs (dash-and-dotted), and standard deviation (solid). For each object the shaded areas display the 90 percent posterior credible intervals. Left column: current year; right column: next year.

one-third than the variance for the current year. A key difference relative to the current-year projections is that disagreement is as important a driver for this increase as the average variance across individuals. In 2021 the average variance declines a bit, but remains very high, while disagreement falls by more than half, but remains quantitatively important. The only two episodes where disagreement plays a non-negligible role are the COVID crisis and the Great Recession.

Figure 10 showed that average subjective uncertainty for GDP growth predictions, as measured by the variance, rose dramatically during the COVID pandemic. Was this rise

Figure 15: Inflation at Risk during COVID



Note: 90th quantile for individual (light gray with crosses) and consensus density forecast (dashed black); the solid black line depicts the average across individuals.

common to all forecasters? Is it robust to different measures of uncertainty? (Recall from previous sections that different measures can result in different answers as density forecasts are sometimes far from Gaussian.) Was it due to compositional effects? Figure 11 addresses these questions.

The top panels of Figure 11 use the IQR as a measure of uncertainty (results for the IQD(10,90) are qualitatively similar). Each panel displays the IQR for individual forecasters (thin gray lines with crosses), the mean of the cross-sectional IQR distribution (solid line; the median is very similar and therefore not reported), and the IQR for the consensus forecast (dashed line). We show individual measures of uncertainty so that one can informally assess the extent to which changes in the composition of the panel affect summary measures of uncertainty such as the mean (Manski, 2018, stresses the extent to which the literature has often ignored compositional changes when discussing the evolution of consensus measures). The bottom panels of Figure 11 provide summary measures of heterogeneity in uncertainty across respondents. Specifically, they display the cross-sectional IQR of different measures of individual uncertainty: the IQR, the IQD(10,90), and the standard deviation. The left and right columns show the results for projections concerning the current and following year, respectively.

The left column of Figure 11 shows that the increase in the uncertainty for the current

year is (almost) universal: the IQR rises for all but one forecaster. The mean of the ranges and the range of the consensus forecast increase approximately by the same amount, highlighting that disagreement, while important in absolute terms, is, in relative terms, swamped quantitatively by the increase in uncertainty. The change in uncertainty for current-year projections differs widely across respondents, however. This is shown by the fact that the cross-sectional IQRs of various measures of uncertainty all rise substantially in 2020 (bottom-left panel). For instance the IQR for both standard deviations and IQR rises from about .5 to 2 in 2020, while the IQR for IQD(10,90) increases from 1 to almost 5.

The situation is very different for the year-ahead forecasts. The increase in volatility in 2020 for the following-year forecast is less noticeable, as observed before. It is also far from homogeneous, with a sizable rise in uncertainty for some forecasters and a decline for others (top-right panel). The increase in disagreement is also evident, as the IQR for the consensus rises more than the average individual IQR. Another difference with the current-year forecasts is that uncertainty remains on average as elevated in 2021 as in 2020. Heterogeneity in uncertainty for year-ahead forecasts also rises in 2020 (bottom-right panel). Partly because the increase in uncertainty is not as pronounced as for current-year projections, the rise in heterogeneity is also less sizable.

Last, we discuss "growth at risk" as measured by the 10th quantiles of the forecast distribution (results for the 5th quantile are similar, if a bit more extreme). The notion of "growth at risk" has become popular following the seminal work of Adrian et al., 2019; see also Kozlowski et al., 2020b, 2019, 2020a for a discussion of the macroeconomic implications of tail risks. In 2020 there is much dispersion in growth at risk across respondents, with the 10th quantiles ranging from -5 to about -18 percent. All quantiles are much lower relative to pre-COVID years, however, indicating that growth at risk in the short run increased dramatically relative to previous years. The picture is quite different for year-ahead forecasts, where growth at risk increases noticeably for some forecasters but declines for others, so that the 10th quantile of the consensus forecast falls in 2020 but the mean declines only slightly.

VI.B Inflation

Turning to density forecasts for GDP deflator inflation, Figure 13 documents that the mean of the consensus distribution, which of course coincides with the average mean projection across forecasters, fell in the period from 2019 to 2020 but then increased noticeably in 2021, especially for the current year. In sharp contrast with the GDP growth projections,

the variance of the consensus forecast for inflation remains well below pre-Great Recession levels for both current-year and year-ahead predictions, although it increased a bit during COVID for current-year forecasts. In fact, much of the increase in the variance of the consensus distribution for current-year projections from 2019 to 2020 is due to an increase in disagreement, as average uncertainty across forecasters rose only modestly. From 2020 to 2021, average uncertainty remained essentially flat for the current year, after accounting for estimation uncertainty, but disagreement declined.

Figure 14 confirms that average inflation uncertainty remains low by historical standards during the COVID period, no matter how it is measured. In fact, for most individuals, uncertainty barely changes from 2019 to 2020. Heterogeneity in uncertainty, as measured by the cross-sectional IQRs also does not rise during the pandemic. Indeed, both average uncertainty and heterogeneity in uncertainty appear to be on a downward trend since the Great Recession—a trend that is barely affected by the COVID period. This trend seems to be driven by high uncertainty forecasters who either changed their view and became more confident about their inflation projections, or disappeared from the sample.

Figure 15 shows that "inflation at risk," as measured by the 90th quantile of the distribution, fell on average from 2019 to 2020 but rose noticeably in 2021 (results for the 95th quantile are similar; to our knowledge, Andrade et al., 2012, coined the term "inflation at risk" and were the first to measure it for SPF forecasters). However, in Q2 this measure was close to or slightly below the levels reached in 2007, before the Great Recession. The dispersion in inflation at risk across forecasters in 2021 is much less pronounced than it was in 2007 or even in 2011, when oil prices rose following the so-called Arab Spring.

VII Conclusions

The past two decades witnessed the emergence of a large number of probabilistic surveys in macroeconomics eliciting predictive probabilities from professional forecasters, financial market participants, consumers, and firms, in the US and other countries. These surveys provide a wealth of information to researchers, who have enthusiastically used them to study several questions, such as the evolution of uncertainty over time.

This chapter reviewed this growing literature, with a particular emphasis on the approaches employed to translate the information provided by forecasters into objects of interest for macroeconomists. It also discussed the substantial inference challenges that this task

entails and presented a novel Bayesian non-parametric approach proposed by Del Negro et al. (2018) that tries to address some of these challenges.

Many questions concerning the proposed approach remain. One question pertains to the robustness with respect to the choice of priors and the base function—e.g., using a mixture of three as opposed to two normals. Another question relates to the robustness of the inference to the survey design and in particular to the location of the bin edges. We performed two experiments using the 2020Q2 real GDP survey. In the first experiment we merged the left open bin $(-\infty, -12]$ with the adjacent closed bin $(-12, -6]$, while in the second we merged the two closed bins $(-12, -6]$ and $(-6, -3]$. This preliminary investigation indicates, not surprisingly, that inference is much sharper when more information is provided—that is, when the bins are not merged. For most of the objects considered, the (5,95) posterior credible intervals from BNP obtained when less information is available (when the bins are merged) include the much narrower (5,95) intervals obtained using all the information from the survey. This suggests that the procedure may adequately reflect the loss of information, although a much more thorough investigation is needed.

Finally, for the time being the BNP approach deals with one survey (one cross-section) and one forecast variable at the time. It would be interesting to extend the approach to a panel context, which would permit joint inference across surveys for any object of interest (e.g., one could test the significance of changes over time for the average uncertainty across forecasters). Similarly, it would be interesting to extend the approach to a multi-variable context, although at the moment we are not aware of any survey that systematically asks probabilistic questions for joint distributions (say, GDP growth and inflation).

As more and more probabilistic surveys are being conducted, it is clear that much more research is needed in order to make progress on the econometric challenges discussed in this chapter—ideally using a variety of diverse approaches, both frequentist and Bayesian. We hope that this review will help spur interest in such research.

References

Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone, "Vulnerable growth," *American Economic Review*, 2019, *109* (4), 1263–89.

Altig, David, Jose Maria Barrero, Nicholas Bloom, Steven J Davis, Brent Meyer, and Nicholas Parker, "Surveying business uncertainty," *Journal of Econometrics*, 2020.

Andrade, Philippe and Hervé Le Bihan, "Inattentive professional forecasters," *Journal of Monetary Economics*, 2013, *60* (8), 967–982.

_ , **Eric Ghysels, and Julien Idier**, "Tails of inflation forecasts and tales of monetary policy," *Available at SSRN 2185958*, 2012.

_ , **Richard K Crump, Stefano Eusepi, and Emanuel Moench**, "Fundamental disagreement," *Journal of Monetary Economics*, 2016, *83*, 106–128.

Armantier, Olivier, Giorgio Topa, Wilbert Van der Klaauw, and Basit Zafar, "An overview of the survey of consumer expectations," *Economic Policy Review*, 2017, (23-2), 51–72.

Bassetti, Federico, Roberto Casarin, and Fabrizio Leisen, "Beta-product dependent Pitman–Yor processes for Bayesian inference," *Journal of Econometrics*, 2014, *180* (1), 49–72.

_ , _ , **and Francesco Ravazzolo**, "Bayesian Nonparametric Calibration and Combination of Predictive Distributions," *Journal of the American Statistical Association*, 2018, *113* (522), 675–685.

Billio, M., R. Casarin, and L. Rossini, "Bayesian Nonparametric Sparse VAR Models," *Journal of Econometrics*, 2019, *212*, 97–115.

_ , _ , **F. Ravazzolo, and H. Van Dijk**, "Time-varying Combinations of Predictive Densities using Nonlinear Filtering," *Journal of Econometrics*, 2013, *177*, 213–232.

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

Bloom, Nicholas, "The impact of uncertainty shocks," *econometrica*, 2009, *77* (3), 623–685.

_ , "Fluctuations in uncertainty," *Journal of Economic Perspectives*, 2014, *28* (2), 153–76.

Boero, G., Jeremy Smith, and Kenneth F. Wallis, "The Measurement and Characteristics of Professional Forecasts' Uncertainty," *Journal of Applied Econometrics*, 2014, *7* (30), 1029–1046.

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

_ , _ , **and** _ , "Uncertainty and disagreement in economic prediction: the Bank of England Survey of External Forecasters," *The Economic Journal*, 2008, *118* (530), 1107–1127.

Bomberger, William A and William J Frazer, "Interest rates, uncertainty and the Livingston data," *The Journal of Finance*, 1981, *36* (3), 661–675.

Bruine De Bruin, Wändi, Charles F Manski, Giorgio Topa, and Wilbert Van Der Klaauw, "Measuring consumer uncertainty about future inflation," *Journal of Applied Econometrics*, 2011, *26* (3), 454–478.

Campbell, Sean D, "Macroeconomic volatility, predictability, and uncertainty in the great moderation: evidence from the Survey of Professional Forecasters," *Journal of Business & Economic Statistics*, 2007, *25* (2), 191–200.

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

Carroll, Christopher D, "Macroeconomic expectations of households and professional forecasters," *the Quarterly Journal of economics*, 2003, *118* (1), 269–298.

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

_ , "Forecast Uncertainty—Ex Ante and Ex Post: US Inflation and Output Growth," *Journal of Business & Economic Statistics*, 2014, *32* (2), 206–216.

_ , "Probability distributions or point predictions? Survey forecasts of US output growth and inflation," *International Journal of Forecasting*, 2014, *30* (1), 99–117.

_ **and Ana Beatriz Galvão**, "Model and survey estimates of the term structure of US macroeconomic uncertainty," *International Journal of Forecasting*, 2017, *33* (3), 591–604.

Clements, Michael P. and David I. Harvey, "Combining probability forecasts," *International Journal of Forecasting*, 2011, *27* (2), 208–223.

ClementsetalCHAPT-inVOLUME, "Surveys of Professionals," in R, ed., *Handbook of Economic Expectations*, Elsevier, 2022.

Conflitti, Cristina, Christine De Mol, and Domenico Giannone, "Optimal combination of survey forecasts," *International Journal of Forecasting*, 2015, *31* (4), 1096–1103.

Cumings-Menon, Ryan, Minchul Shin, and Keith Sill, "Measuring Disagreement in Probabilistic and Density Forecasts," Technical Report, Federal Reserve Bank of Philadelphia 2021.

D'Amico, Stefania and Athanasios Orphanides, "Uncertainty and disagreement in economic forecasting," Technical Report, Board of Governors of the Federal Reserve System (US) 2008.

Daniel, Kent and David Hirshleifer, "Overconfident investors, predictable returns, and excessive trading," *Journal of Economic Perspectives*, 2015, *29* (4), 61–88.

DeGroot, M. H., A. P. Dawid, and J. Mortera, "Coherent combination of experts' opinions," *Test*, 1995, *4*, 263–313.

DeGroot, Morris H. and Julia Mortera, "Optimal Linear Opinion Pools," *Management Science*, 1991, *37* (5), 546–558.

Del Negro, Marco, Raiden B. Hasegawa, and Frank Schorfheide, "Dynamic prediction pools: An investigation of financial frictions and forecasting performance," *Journal of Econometrics*, 2016, *192* (2), 391–405.

_ , Roberto Casarin, and Federico Bassetti, "A Bayesian Approach for Inference on Probabilistic Surveys," Technical Report, Mimeo 2018.

Dominitz, Jeff and Charles F Manski, "Eliciting student expectations of the returns to schooling," *Journal of Human resources*, 1996, pp. 1–26.

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

_ , _ , and _ , "Assessing the temporal variation of macroeconomic forecasts by a panel of changing composition," *Journal of Applied Econometrics*, 2011, *26* (7), 1059–1078.

Ferguson, T. S., "A Bayesian analysis of some nonparametric problems," *Annals of Statistics*, 1973, *1*, 209–230.

Ganics, Gergely, Barbara Rossi, and Tatevik Sekhposyan, "From fixed-event to fixed-horizon density forecasts: obtaining measures of multi-horizon uncertainty from survey density forecasts," 2020.

Garcia, Juan A and Andrés Manzanares, "What can probability forecasts tell us about inflation risks?," 2007.

Genest, Christian and James V Zidek, "Combining probability distributions: A critique and an annotated bibliography," *Statistical Science*, 1986, *1* (1), 114–135.

Genre, Véronique, Geoff Kenny, Aidan Meyler, and Allan Timmermann, "Combining expert forecasts: Can anything beat the simple average?," *International Journal of Forecasting*, 2013, *29* (1), 108–121.

Geweke, J. and G. Amisano, "Optimal Prediction Pools," *Journal of Econometrics*, 2011, *164*, 130–141.

Ghosh, J. K. and R. V. Ramamoorthi, *Bayesian nonparametrics* Springer Series in Statistics, Springer-Verlag, New York, 2003.

Giordani, Paolo and Paul Soderlind, "Inflation forecast uncertainty," *European Economic Review*, 2003, *47*, 1037–1059.

Giustinelli, Pamela, Charles F Manski, and Francesca Molinari, "Tail and center rounding of probabilistic expectations in the health and retirement study," *Journal of Econometrics*, 2020.

Griffin, J. and M. Kalli, "Bayesian Nonparametric Vector Autoregressive Models," *Journal of Econometrics*, 2018, *203*, 267–282.

Griffin, J. E. and M. F. J. Steel, "Stick-breaking Autoregressive Processes," *Journal of Econometrics*, 2011, *162*, 383–396.

Hall, Stephen G and James Mitchell, "Combining density forecasts," *International Journal of Forecasting*, 2007, *23* (1), 1–13.

Hirano, K., "Semiparametric Bayesian Inference in Autoregressive Panel Data Models," *Econometrica*, 2002, *70*, 781–799.

Jurado, Kyle, Sydney C Ludvigson, and Serena Ng, "Measuring uncertainty," *American Economic Review*, 2015, *105* (3), 1177–1216.

Kapetanios, G., J. Mitchell, S. Price, and N. Fawcett, "Generalised Density Forecast Combinations," *Journal of Econometrics*, 2015, *188*, 150–165.

Karabatsos, George and Stephen G Walker, "Coherent psychometric modelling with Bayesian nonparametrics," *British Journal of Mathematical and Statistical Psychology*, 2009, *62* (1), 1–20.

Kenny, Geoff, Thomas Kostka, and Federico Masera, "How informative are the subjective density forecasts of macroeconomists?," *Journal of Forecasting*, 2014, *33* (3), 163–185.

Kozlowski, Julian, Laura Veldkamp, and Venky Venkateswaran, "The tail that keeps the riskless rate low," *NBER Macroeconomics Annual*, 2019, *33* (1), 253–283.

_ , _ , and _ , "Scarring body and mind: the long-term belief-scarring effects of Covid-19," Technical Report, National Bureau of Economic Research 2020.

_ , _ , and _ , "The tail that wags the economy: Beliefs and persistent stagnation," *Journal of Political Economy*, 2020, *128* (8), 2839–2879.

Lahiri, Kajal and Fushang Liu, "Modelling multi-period inflation uncertainty using a panel of density forecasts," *Journal of Applied Econometrics*, 2006, *21* (8), 1199–1219.

Li, Yuelin, Elizabeth Schofield, and Mithat Gönen, "A tutorial on Dirichlet process mixture modeling," *Journal of mathematical psychology*, 2019, *91*, 128–144.

Liu, Yang and Xuguang Simon Sheng, "The measurement and transmission of macroeconomic uncertainty: Evidence from the U.S. and BRIC countries," *International Journal of Forecasting*, 2019, *35* (3), 967–979.

Malmendier, Ulrike and Timothy Taylor, "On the verges of overconfidence," *Journal of Economic Perspectives*, 2015, *29* (4), 3–8.

Mankiw, N Gregory, Ricardo Reis, and Justin Wolfers, "Disagreement about inflation expectations," *NBER macroeconomics annual*, 2003, *18*, 209–248.

Manski, Charles F, "Measuring expectations," *Econometrica*, 2004, *72* (5), 1329–1376.

_ , "Interpreting and combining heterogeneous survey forecasts," in M. P. Clements and D. F. Hendry, eds., *Oxford handbook of economic forecasting*, Vol. 85, Oxford University Press, 2011, pp. 457–472.

_ , "Survey measurement of probabilistic macroeconomic expectations: progress and promise," *NBER Macroeconomics Annual*, 2018, *32* (1), 411–471.

_ **and Francesca Molinari**, "Rounding probabilistic expectations in surveys," *Journal of Business & Economic Statistics*, 2010, *28* (2), 219–231.

Manzan, Sebastiano, "Are professional forecasters Bayesian?," *Journal of Economic Dynamics and Control*, 2021, *123*, 104045.

McAlinn, Kenichiro and Mike West, "Dynamic Bayesian predictive synthesis in time series forecasting," *Journal of econometrics*, 2019, *210* (1), 155–169.

Mirkov, Nikola and Andreas Steinhauer, "Asymmetry of Individual and Aggregate Inflation Expectations: A Survey," *The Manchester School*, 2018, *86* (4), 446–467.

Navarro, Daniel J, Thomas L Griffiths, Mark Steyvers, and Michael D Lee, "Modeling individual differences using Dirichlet processes," *Journal of mathematical Psychology*, 2006, *50* (2), 101–122.

Patton, Andrew J and Allan Timmermann, "Why do forecasters disagree? Lessons from the term structure of cross-sectional dispersion," *Journal of Monetary Economics*, 2010, *57* (7), 803–820.

_ **and** _ , "Predictability of output growth and inflation: A multi-horizon survey approach," *Journal of Business & Economic Statistics*, 2011, *29* (3), 397–410.

Pitman, J., *Combinatorial Stochastic Processes*, Vol. 1875, Springer-Verlag, 2006.

Ranjan, Roopesh and Tilmann Gneiting, "Combining probability forecasts," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2010, *72* (1), 71–91.

Rich, Robert and Joseph Tracy, "A closer look at the behavior of uncertainty and disagreement: Micro evidence from the euro area," *Journal of Money, Credit and Banking*, 2021, *53* (1), 233–253.

Scealy, J. L. and A. H. Welsh, "Regression for compositional data by using distributions defined on the hypersphere," *Journal of the Royal Statistical Society, Series B*, 2011, *73* (3), 351–375.

Shoja, Mehdi and Ehsan S. Soofi, "Uncertainty, information, and disagreement of economic forecasters," *Econometric Reviews*, 2017, *36* (6-9), 796–817.

Stark, Tom, "SPF panelists' forecasting methods: A note on the aggregate results of a November 2009 special survey," *Federal Reserve Bank of Philadelphia*, 2013.

Stone, M, "The opinion pool," *Annals of Mathematical Statistics*, 1961, *32*, 1339–1342.

Timmermann, Allan, "Forecast Combinations," in Graham Elliott, Clive Granger, and Allan Timmermann, eds., *Handbook of Economic Forecasting*, Vol. 1 of *Handbooks in Economics 24*, North Holland, Amsterdam, 2006, pp. 135–196.

Wachtel, Paul, "Survey measures of expected inflation and their potential usefulness," Technical Report, National Bureau of Economic Research 1977.

Zadora, G., T. Neocleous, and Aitken C., "A Two-Level Model for Evidence Evaluation in the Presence of Zeros," *Journal of Forensic Sciences*, 2010, *55* (2), 371–384.

Zarnowitz, V. and L.A. Lambros, "Consensus and Uncertainty in Economic Prediction," *Journal of Political Economy*, 1987, (95), 591 – 621.

Zarnowitz, Victor, "Front matter," An Appraisal of Short-term Economic Forecasts"," in "An Appraisal of Short-Term Economic Forecasts," Nber, 1967, pp. 14–0.