STAFF
REPORTS

NO. 1025
JULY 2022

REVISED
AUGUST 2024

A Bayesian Approach for Inference on Probabilistic Surveys

Federico Bassetti | Roberto Casarin | Marco Del Negro

FEDERAL RESERVE BANK *of* NEW YORK

A Bayesian Approach for Inference on Probabilistic Surveys

Federico Bassetti, Roberto Casarin, and Marco Del Negro

Federal Reserve Bank of New York Staff Reports, no. 1025
July 2022; Revised August 2024
JEL classification: C11, C14, C53, C82, E31, E32, E37

Abstract

We propose a nonparametric Bayesian approach for conducting inference on probabilistic surveys. We use this approach to study whether U.S. Survey of Professional Forecasters density projections for output growth and inflation from 1982 to 2022 are consistent with the noisy rational expectations hypothesis. We find that, in contrast to theory, for horizons close to two years there is no relationship whatsoever between subjective uncertainty and forecast accuracy for output growth density projections, both across forecasters and over time, and only a mild relationship for inflation projections. As the horizon shortens, the relationship becomes one-to-one as theory predicts.

Key words: Bayesian nonparametrics, probabilistic surveys, noisy rational expectations.

Del Negro: Federal Reserve Bank of New York (email: marco.delnegro@ny.frb.org). Casarin: Ca' Foscari University of Venice (email: r.casarin@unive.it). Bassetti: Polytechnic University of Milan (email: federico.bassetti@polimi.it). This project was begun with Francesco Ravazzolo, whom the authors thank for many useful conversations. They also thank Alissa Johnson for research assistance; the participants of several seminars, workshops, and conferences for their comments; and Michael Clements, Elmar Mertens, and Christopher Sims for their detailed feedback.

I Introduction

The pioneering work of Manski (2004) highlighted the benefits of probabilistic surveys compared to surveys that only ask respondents for their point projections: probabilistic surveys simply provide a wealth of information that is not included in point projections.[1] As Potter (2016) writes, "In a world characterized by pervasive uncertainty, density forecasts provide a comprehensive representation of respondents' views about possible future outcomes for the variables of interest." Given the respondents' density forecasts, the econometrician can compute numerous objects of interest, such as the mean, median, variance, skewness, interquantile range, *et cetera*.

However, survey respondents do not provide us with density forecasts. For most surveys concerning continuous variables, they only provide the percent chance that the variable of interest (e.g., inflation over the next year) would fall within different prespecified contiguous ranges or bins. That is, the information we have consists in the integral of the forecast density over these bins or, equivalently, in a few points of the cumulative density function (CDF). In order to extract many quantities of interest, and in particular measures of uncertainty, standard practice postulates a parametric form for the forecast distribution and computes its parameters by minimizing the distance between the observed CDF points and those implied by the assumed distribution, which is often either a step-wise uniform (Zarnowitz and Lambros, 1987), a Gaussian (Giordani and Soderlind, 2003), or a generalized beta distribution (Engelberg et al., 2009).[2]

In this paper, we propose a Bayesian nonparametric approach for estimating the survey respondents' forecast densities.[3] The approach starts by making parametric assumptions

[1] Indeed, a number of recent surveys, including the Federal Reserve Bank of New York's Survey of Consumer Expectations and Survey of Primary Dealers and Market Participants, the Canadian Survey of Consumer Expectations, the Atlanta Feds Survey of Business Uncertainty and Business Inflation Expectations, the ECBs Survey of Professional Forecasters, and the the Bank of Englands Survey of External Forecasters, elicit probabilistic questions.

[2] For a few quantities of interest, such as the median, one can compute nonparametric bounds as in Engelberg et al. (2009). Computing nonparametric bounds is harder for measures of uncertainty, including interquantile ranges, once one acknowledges the issue of noise/rounding. In any case, the parametric approach described in the text is the one most commonly used in the literature, and by the surveys mentioned in footnote 1, when reporting measures of uncertainty.

[3] In economics, the Bayesian nonparametric approach so far has applied to the analysis of treatment effects (Chib and Hamilton, 2002), autoregressive panel data (Hirano, 2002; Gu and Koenker, 2017; Liu, 2023), time series (Bassetti et al., 2014), stochastic production frontiers models (Griffin and Steel, 2004), unemployment

on the mapping between the predictive distribution of forecasters and the bin probabilities they report, where this mapping explicitly allows for the introduction of noise in the reporting (due to rounding toward zero, for example). We then relax this parametric model by embedding it into the more general Bayesian nonparametric framework, thereby amending the potential misspecification associated with the parametric assumptions. This is because, loosely speaking, Bayesian nonparametrics replaces any parametric model with a potentially infinite mixture of such models, attaining more flexibility while at the same time using the information from the cross-section of forecasters to estimate the parameters of the mixture components. Intuitively, each mixture component corresponds to a forecaster type (e.g., low/high mean, low/high variance, low/high noise, *et cetera*, and combinations thereof). As long as the number of types grows more slowly than the number of forecasters, there is enough information to estimate the parameters corresponding to each type.

Our approach differs from existing methods in a few important dimensions. First, inference conducted using a specific parametric distribution can be naturally sensitive to the choice of the distribution. The nonparametric nature of our approach provides some robustness to misspecification regarding these parametric assumptions. Second, our approach conducts inference jointly across survey respondents, that is, using the entire cross-section instead of being applied to each respondent separately. As mentioned above, this joint inference allows for partial information-pooling across forecasters, thereby improving inference precision and making it possible to obtain some consistency results when the number of forecasters grows to infinity. Last, the approach allows for full-fledged inference regarding the mapping between data and objects of interest, in the sense that it generates a posterior probability for these objects. While current approaches provide point estimates for, say, measures of the scale of the predictive densities like the variance, they do not supply any assessment of the uncertainty surrounding these estimates, which is often large given the limited information provided by survey responses.

We use this approach to address the question of whether U.S. Survey of Professional Forecasters (SPF) density forecasts are consistent with the noisy rational expectations (RE) hypothesis (see, for instance, Coibion and Gorodnichenko, 2012, 2015) using data from 1982 to 2022. According to this hypothesis, forecasters receive both public and private signals about the state of the economy. The precision of forecasters' signals, both public and private, ought to be reflected in equal measure in their density forecasts and, under RE, in their ex-

duration (Burda et al., 2015), and finance (Griffin, 2011, and Jensen and Maheu, 2010). Griffin et al. (2011) provide an intuitive description of the approach and a survey of this literature.

post forecast accuracy, both in the cross-section and over time. For example, if the economy becomes more uncertain, this should be reflected in both higher subjective uncertainty *and* worse *ex-post* forecast errors. In fact, we find that for horizons close to two years there is no relationship whatsoever between subjective uncertainty and *ex-post* forecast accuracy for output growth density projections, and only a mild relationship for inflation projections. As the horizon shortens, the relationship becomes close to one-to-one, in accordance with the theory. These findings, which hold when controlling for both forecaster and time-fixed effects, suggest that forecasters do not correctly anticipate periods of macroeconomic uncertainty, except for very short horizons. We also find that forecasters tend to be overconfident for long horizons, but underconfident for short horizons, although the RE benchmark is never rejected for inflation.

The outline of the paper is as follows. Section II presents the inference problem, briefly describes current approaches, and formally discusses our Bayesian nonparametric approach. Section III first provides a few examples of how our approach differs from current practice and then discusses the relationship between subjective uncertainty and forecast accuracy. Section IV concludes, pointing out some of the limitations of the analysis and discussing avenues for further research.

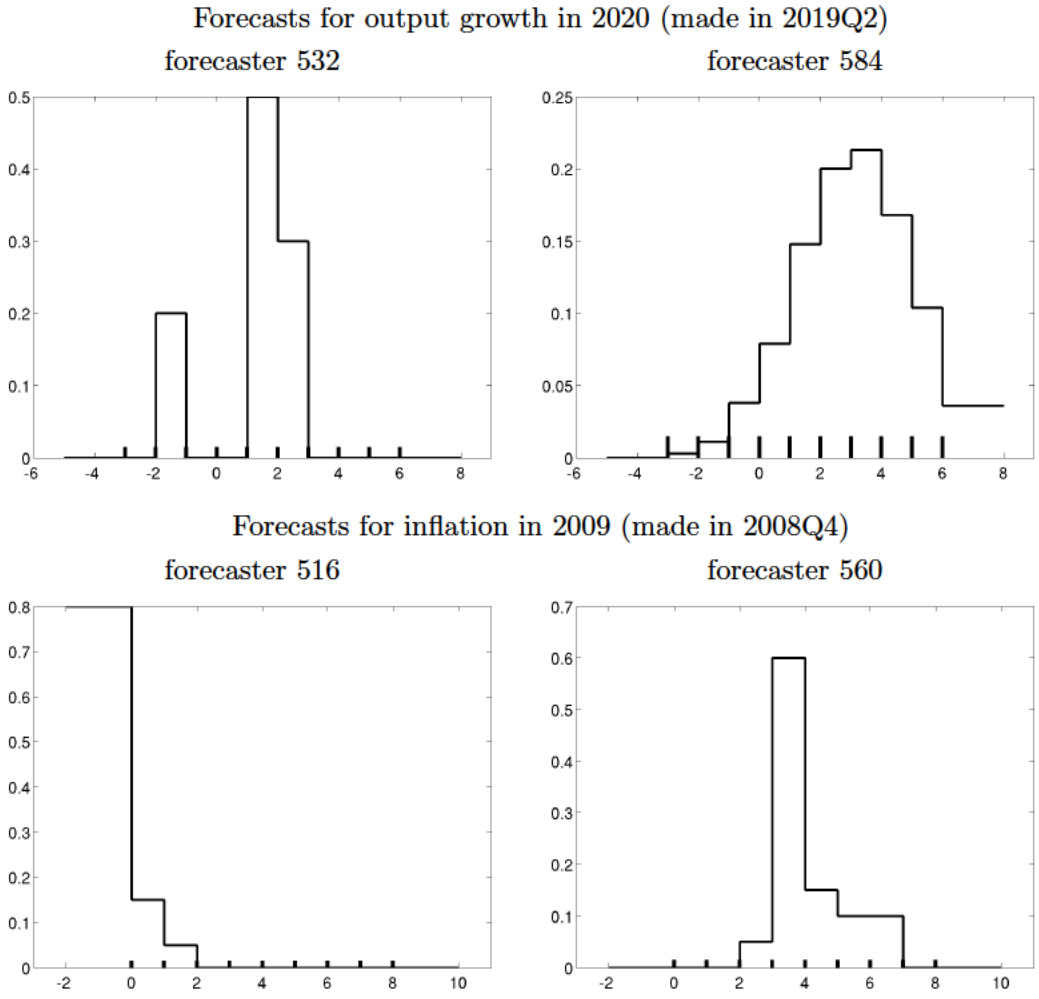
II Inference for Probabilistic Surveys

In this section we start by providing a short introduction to probabilistic survey data, focusing on those features that are relevant for this analysis, and in the process describe the SPF data used in our application. Then we briefly discuss the approaches used so far for translating the information provided by the respondents into subjective predictive distributions and objects of interest, such as measures of uncertainty. The rest of the section is devoted to the description of our approach.

II.A The inference problem and current approaches

Probabilistic forecasts such as those elicited by the Philadelphia Fed as part of the SPF take the form of probabilities assigned to bins: the percent chance that the variable of interest,

Figure 1: Probability forecasts for selected examples



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

for example, inflation, falls within different contiguous ranges, where these ranges are pre-specified by the survey designer.[4] Formally, for each forecaster $i = 1, ..., n$ the data consist of a vector of probabilities $\mathbf{z}_i = (z_{i1}, ..., z_{iJ})$, with $z_{ij} \geq 0$ and $\sum_{j=1}^{J} z_{ij} = 1$, measuring the predictive likelihood that the continuous variable y falls within the respective bin. Reflecting the fact that the bins are mutually exclusive and contiguous, and generally cover the entire real line, we denote the bins by $(y_{j-1}, y_j]$, $j = 1, ..., J$, with $y_0 < y_1 < ... < y_J$, where y_0 and y_J are equal to $-\infty$ (left open bin) and $+\infty$ (right open bin), respectively. Figure A-3

[4]Some recent surveys, such as the Atlanta Fed's Survey of Business Uncertainty, only specify the number of bins and let the respondents determine their boundaries.

in the appendix displays the evolution of the bin ranges from 1982, the beginning of our sample, until the end in 2022, for both output growth and inflation. The figure shows that bins were changed in 1992, 2009, and 2020 for output growth surveys and in 1985, 1992, and 2014 for inflation surveys. The SPF is conducted at a quarterly frequency (answers are collected in the middle of each quarter, right after GDP figures for the previous quarter have been released) and asks about probabilistic predictions for the current and the following year's year-over-year growth rates in real output (GDP) and the price level, as measured by the GDP deflator. Stark (2013) describes in detail the features of the SPF survey, and the Philadelphia Fed's site provides a manual for interpreting the data.[5]

Figure 1 provides a few examples that illustrate a number of common features of the SPF data. The top two panels show the probabilistic forecasts for output growth in 2020 made in 2019Q2 by respondents 532 and 584, while the bottom two panels show the forecasts for inflation in 2009 made in 2008Q4 by respondents 516 and 560 (respondents are anonymous). The probabilities $\mathbf{z}_i$'s are displayed as histograms, while the black ticks on the horizontal axis mark the boundaries of the bins.

A first feature that emerges from Figure 1 is that probabilistic forecasts are very heterogeneous. For each row the respondents are forecasting the same object, and yet their probabilistic predictions are very different. A second feature is that forecasters often assign zero probability to some if not most bins. Forecaster 532, for instance, places zero probability on output growth being between -1 and 1 percent, but positive probability on output being between -2 and -1 percent and between 1 and 3 percent. The econometrician could interpret this information literally, or as an indication that this respondent has a bimodal forecast distribution with some probability on a recession, a larger probability on an expansion, and a very small likelihood of in-between outcomes. Other forecasters, such as respondent 584, place positive mass on almost all bins. A third feature of the data is that almost all probabilities in Figure 1 are round numbers, with responses for forecaster 584 being, again, the only exception. A fourth feature is that forecasters do place mass on open bins and sometimes, as is the case for the respondent who in 2008 was fearing deflation in 2009, most of the mass. Figures A-6 and A-7 in the appendix show for each output growth and inflation survey the

[5]Figure A-4 in the appendix displays the number of respondents n for output growth surveys conducted in Q1, Q2, Q3, and Q4 of each year (the numbers for inflation are essentially the same). n is about 35 in the early 1980s and then drops steadily over time until 1992, when the Philadelphia Fed begins to manage the survey; n hovers around 35 until the mid-2000s and then starts to increase, reaching a peak of about 50 during the Great Recession; it declines steadily thereafter and is about 30 in 2022. Figure A-5 shows survey participation by respondent and provides a visual description of the panel's composition.

percentage of respondents placing positive probability on either one open bin or both. These percentages are as high as 70 percent for output and 90 percent for inflation before 1992, when the bins were changed, but are about 20 percent on average, with peaks of 40 percent or higher, even after 1992. Finally, many of these predictive densities appear asymmetric. These examples display a left skew for output and, at least for forecaster 560, a right skew for inflation.

The econometricians problem is how to use the information 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? Is there skewness in their predictive densities? The approach predominantly used so far in the literature concerning macroeconomic surveys has been to implicitly or explicitly assume that each forecaster i assigns the bin probabilities $\mathbf{z}_i$ using a given predictive probability distribution $F_i(y)$. The econometrician then needs to estimate $F_i(y)$ from the data $\mathbf{z}_i$ and use it to answer the questions of interest. Most 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 $Z_{ij} = z_{i1} + \cdots + z_{ij}$ $j = 1, \ldots, J$, $i = 1, ..., n$ 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 generalized beta (Engelberg et al., 2009), and a skew-t distribution (e.g., Ganics et al., 2020). The generalized beta assumption is arguably the most popular approach. The parameters of each distribution are usually estimated using nonlinear least squares, respondent by respondent; that is, $F_i(y) = F(y|\hat{\boldsymbol{\theta}}_i)$, where

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

These approaches arguably have some limitations that are well understood in the literature (see Clements et al., 2023). First, the assumed parametric distribution may be misspecified—it may not fit the individual responses well. Moreover, since the width of the bins can be large (as is obviously the case when the respondent places probability on open bins), even if the distributions fit the Z_{ij}'s, the inference results on moments and quantiles can be sensitive to the distributional assumption. Second, and related, existing approaches ignore inference uncertainty, even that concerning $\boldsymbol{\theta}_i$ for a given parametric assumption, let

alone the uncertainty about the shape of $F_i(\cdot)$. This omission implies that confidence bands and hypothesis testing procedures are usually not derived.[6] Third, bounded distributions such as the beta or the mixture of uniforms take literally the z_{ij} that are zero—they place no probability mass on bins where the respondents place no mass. More broadly, the approach outlined in expression (1) does not directly address the issue of rounding: it solves the minimization problem taking all the Z_{ij}'s literally even though the respondent may be reporting approximate probabilities (Dominitz and Manski, 1996; D'Amico and Orphanides, 2008; Boero et al., 2008, 2014; Engelberg et al., 2009; Manski and Molinari, 2010; Manski, 2011; Giustinelli et al., 2020; Glas and Hartmann, 2022, among others, discuss the issue of rounding; Binder, 2017, uses rounding to measure uncertainty).[7]

In the following two sections, we propose a Bayesian model that attempts to overcome some of these limitations. We first introduce a *parametric* model for the data. This model follows the literature in assuming that each forecaster uses a specific predictive distribution $F(\cdot)$ to assign probabilities $\boldsymbol{\nu}$ to the bins, but differs from the literature in that it states that the data $\mathbf{z}$ are noisy versions of the $\boldsymbol{\nu}$'s, where again the noise follows a parametric distribution. We then relax this parametric model by embedding it into the more general Bayesian *nonparametric* approach, thereby amending the potential misspecification associated with the parametric assumptions.

II.B A parametric model

We assume that the probability vector $\mathbf{z}_i$ reported by forecaster i is a noise-ridden measurements of an unobserved vector of subjective probabilities over the J bins $\boldsymbol{\nu}_i = (\nu_{i1}, ..., \nu_{iJ})$, with $\nu_{ij} \geq 0$ and $\nu_{i1} + ... + \nu_{iJ} = 1$. These bin probabilities ν_{ij} are computed using forecaster

[6] Researchers have of course understood the presence of an inference issue especially when the information provided by the respondent is very limited. The proposed solutions generally amount to either choosing less heavily parameterized distributions or discarding the respondent: Clements (2010), for instance, simply discards respondents with fewer than three bins, while Engelberg et al. (2009) use a triangle distribution in these cases. Liu and Sheng (2019) propose a maximum-likelihood estimation approach in order to account for parameter uncertainty for given parametric assumptions.

[7] Manski and Molinari (2010) and Giustinelli et al. (2020) address 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. The inferential approach pursued by these researchers is very different from the one followed by much of the literature and addresses different questions.

i's subjective probability distribution $F_i(\cdot)$:

$$\nu_{ij} = \nu_j(\boldsymbol{\theta}_i) = F(y_j|\boldsymbol{\theta}_i) - F(y_{j-1}|\boldsymbol{\theta}_i),\ j = 1, \ldots, J, \tag{2}$$

where $\boldsymbol{\theta}_i \in \Theta$ is the vector of all estimated parameters which includes those describing the CDF $F_i(\cdot) = F(\cdot|\boldsymbol{\theta}_i)$. In our application, the subjective distribution $F(\cdot|\boldsymbol{\theta})$ is a mixture of two Gaussian distributions, that is

$$F(y|\boldsymbol{\theta}) = (1-\omega)\Phi(y|\mu, \sigma_1^2) + \omega\Phi(y|\mu + \mu_\delta, \sigma_2^2), \tag{3}$$

but the approach accommodates any other choice for $F(\cdot|\boldsymbol{\theta})$.

The probability distribution $h(\cdot)$ captures the noise in the mapping between $\boldsymbol{\nu}_i$ and $\mathbf{z}_i$, where this noise can be due to approximations or rounding in reporting:

$$\mathbf{z}_i = (z_{i1}, \ldots, z_{iJ}) \overset{ind}{\sim} h(\mathbf{z}_i|\boldsymbol{\nu}(\boldsymbol{\theta}_i), \boldsymbol{\theta}_i). \tag{4}$$

In choosing $h(\cdot)$, one needs to account for the fact that the elements of $\mathbf{z}_i$ are positive and sum up to one ($\mathbf{z}_i$ belongs to the simplex). A convenient choice for a distribution on the simplex is the Dirichlet distribution (Shoja and Soofi, 2017, is another example of using the Dirichlet distribution to model the noise). A drawback of this distribution is that it assigns zero probability to $\mathbf{z}_i$'s that have some elements equal to zero, when, in fact, for the vast majority of forecasters some z_{ij}'s are zero. To specify $h(\cdot)$, we then follow Zadora et al. (2010) and use a distribution that allows for values of the random vector on the boundary of the simplex.

In order to describe this distribution, it is useful to introduce the equivalent representation of $\mathbf{z}_i$ given by the couple $(\mathbf{z}_{\boldsymbol{\xi}i}, \boldsymbol{\xi}_i)$, where $\boldsymbol{\xi}_i = (\xi_{i1}, \ldots, \xi_{iJ})$ with $\xi_{ij} = 1$ if and only if $z_{ij} = 0$ and $\xi_{ij} = 0$ otherwise, and $\mathbf{z}_{\boldsymbol{\xi}i}$ is the set of strictly positive z_{ij}'s. Using these definitions we can write the $h(\cdot)$ distribution as

$$h(\mathbf{z}_i|\boldsymbol{\nu}(\boldsymbol{\theta}_i), \boldsymbol{\theta}_i) = \frac{1}{c(\boldsymbol{\theta}_i)} \prod_{j=1}^{J} \varrho_j(\boldsymbol{\theta}_i)^{\xi_{ij}} (1 - \varrho_j(\boldsymbol{\theta}_i))^{1-\xi_{ij}} \mathcal{D}ir(\mathbf{z}_{\boldsymbol{\xi}i}|\boldsymbol{\nu}(\boldsymbol{\theta}_i), \boldsymbol{\theta}_i), \tag{5}$$

where $\boldsymbol{\varrho}(\boldsymbol{\theta}_i) = (\varrho_1(\boldsymbol{\theta}_i), ..., \varrho_J(\boldsymbol{\theta}_i))$ are the probabilities that a forecaster will report zero probability on bin 1 through J, and $c(\boldsymbol{\theta}_i) = 1 - (\varrho_1(\boldsymbol{\theta}_i) \cdot \ldots \cdot \varrho_J(\boldsymbol{\theta}_i))$ is a normalizing constant. $\mathcal{D}ir(\mathbf{z}_{\boldsymbol{\xi}i}|\boldsymbol{\nu}(\boldsymbol{\theta}_i), \boldsymbol{\theta}_i)$ is the standard Dirichlet distribution defined on the nonzero elements of $\mathbf{z}_i$:

$$\mathcal{D}ir(\mathbf{z}_{\boldsymbol{\xi}i}|\boldsymbol{\nu}(\boldsymbol{\theta}_i), \boldsymbol{\theta}_i) = \frac{\Gamma\left(\sum_{j:z_{ij}>0} \phi_i \nu_j(\boldsymbol{\theta}_i)\right)}{\prod_{j:z_{ij}>0} \Gamma(\phi_i \nu_j(\boldsymbol{\theta}_i))} \prod_{j:z_{ij}>0} z_{ij}^{\phi_i \nu_j(\boldsymbol{\theta}_i) - 1}, \tag{6}$$

where $\phi_i\kappa_i$, $\kappa_i = \sum_{j:z_{ij}>0} \nu_j(\boldsymbol{\theta}_i)$ is the rescaled precision and ϕ_i a parameter. The renormalized weights $\nu_j(\boldsymbol{\theta}_i)/\kappa_i$, j such that $z_{ij} > 0$ take into account the fact that if forecasters decide to report zero probability for one or more bins, they need to adjust the probabilities associated with the other bins so that they still sum up to one.[8]

The probability of reporting zero mass in the j-th bin is modeled as $\varrho_j(\boldsymbol{\theta}_i) = \varrho(\nu_j(\boldsymbol{\theta}_i), \epsilon_i)$, where the parameter ϵ_i measures the sensitivity of $\varrho(\nu, \epsilon)$ to ν and the function $\varrho(\nu, \epsilon)$ is decreasing in ν, such that $\varrho \to 1$ for $\nu \to 0$ and $\varrho \to 0$ for $\nu \to 1$. In our application we use the function

$$\varrho(\nu, \epsilon) = \int_0^{\epsilon} b(x|\nu, r)dx, \tag{7}$$

where $b(x|\nu, r)$ is the PDF of a beta distribution with mean ν and precision r, which we fix at 100. The vector of parameters of $h(\mathbf{z}_i|\boldsymbol{\nu}(\boldsymbol{\theta}_i), \boldsymbol{\theta}_i)$—which in the remainder of this section we will refer to as $h(\mathbf{z}_i|\boldsymbol{\theta}_i)$ for brevity—is therefore $\boldsymbol{\theta}_i = (\mu_i, \mu_{\delta,i}, \sigma_{1,i}, \sigma_{2,i}, \omega_i, \phi_i, \epsilon_i)$.

Some of the parametric assumptions outlined above are less palatable than others. For instance, the assumption that the noise around the nonzero z_{ij}'s takes the form of a Dirichlet distribution is at odds with the observation on the prevalence of rounding. Even when the parametric assumption may be more palatable, it can still be incorrect. Embedding these parametric assumptions into a (more flexible) nonparametric model arguably protects us, at least to some extent, from misspecification. We describe this approach in the next section.

II.C A Bayesian nonparametric model

The Bayesian nonparametric hierarchical model works as follows. We assume that the parameter vector $\boldsymbol{\theta}_i$ is sampled from a mixture of forecaster types (where the types are, for example, low versus high uncertainty, low versus high mean, left versus right-skewed, low versus high reporting noise, or combinations of all the above). For now, imagine that the number of types K is finite. At the first stage of the hierarchy, $\boldsymbol{\theta}_i$ is drawn from

$$\boldsymbol{\theta}_i \overset{iid}{\sim} \begin{cases} \boldsymbol{\theta}_1^* & \text{with probability } w_1 \\ \vdots & \\ \boldsymbol{\theta}_K^* & \text{with probability } w_K \end{cases} \tag{8}$$

[8] Recall that in the Dirichlet distribution the means are $\mathbb{E}(z_{ij}|\boldsymbol{\xi}_i) = \dfrac{\nu_j(\boldsymbol{\theta}_i)}{\sum_{j:z_{ij}>0} \nu_j(\boldsymbol{\theta}_i)}$, and the variances $\mathbb{V}(z_{ij}|\boldsymbol{\xi}_i) = \dfrac{\nu_j(\boldsymbol{\theta}_i)(\kappa_i - \nu_j(\boldsymbol{\theta}_i))}{\kappa_i^2(\phi_i \sum_{j:z_{ij}>0} \nu_j(\boldsymbol{\theta}_i) + 1)}$ go to zero with $\phi_i \to \infty$.

where the w_k's represent probability weights, and $\boldsymbol{\theta}_k^*$ characterize the types, also referred to as atoms. At the second stage a prior distribution is assumed for the weights, and the types are sampled from a common distribution $\boldsymbol{\theta}_k^* \overset{iid}{\sim} G_0$, called a base measure, which can be viewed as the probability distribution generating the types.

Now let the number of types K go to infinity. When this happens, expression (8) is replaced by the discrete random measure

$$G(\boldsymbol{\theta}) = \sum_{k=1}^{\infty} w_k \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_k^*), \tag{9}$$

where $\delta(x)$ denotes a point mass distribution located at 0, the atoms $\boldsymbol{\theta}_k^*$ are drawn from G_0 as before, and the random weights w_k are generated by the so-called stick-breaking construction

$$w_k = v_k \prod_{l=1}^{k-1} (1 - v_l), \tag{10}$$

with the v_l's being i.i.d. random variables from a beta distribution $\mathcal{B}e(1, \psi_0)$ (see Pitman, 2006). The random measure G is a Dirichlet process $\mathcal{DP}(\psi_0, G_0)$ (Ferguson, 1973) and our hierarchical model is a Dirichlet process prior $\boldsymbol{\theta}_i \overset{iid}{\sim} G, \quad G \sim \mathcal{DP}(\psi_0, G_0)$. The precision parameter ψ_0 determines how uneven the weights are in the stick-breaking representation, that is, how many different mixture components are used to fit the respondents: when $\psi_0 \to 0$ all forecasters are assumed to be of the same type ($w_1 \to 1$), whereas when $\psi_0 \to +\infty$ the inference is done forecaster by forecaster (using the same prior). Outside of this latter limiting case, the Dirichlet process prior generates *a priori* dependence among the forecaster-specific parameters $\boldsymbol{\theta}_i$'s via the formation of clusters of forecasters of the same type.[9]

Expression (9) implies that our model has the infinite mixture representation (Sethuraman, 1994)

$$\mathbf{z}_i | G \overset{iid}{\sim} \int h(\mathbf{z}|\boldsymbol{\theta}) G(d\boldsymbol{\theta}) = \sum_{k=1}^{\infty} w_k h(\mathbf{z}|\boldsymbol{\theta}_k^*). \tag{11}$$

Each forecaster is therefore modeled *a priori* as a potentially infinite mixture of types, identical across forecasters (whence the *i.i.d.*), each described by the parametric distribution

[9] As shown in Pitman (2006), the predictive distribution of $\boldsymbol{\theta}_{i+1}$ conditional on $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_i)$ can be represented as a Polya's urn process $\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_i \sim \frac{\psi_0}{\psi_0 + i} G_0(\boldsymbol{\theta}_{i+1}) + \frac{1}{\psi_0 + i} \sum_{k=1}^{i} \delta(\boldsymbol{\theta}_k - \boldsymbol{\theta}_{i+1})$. With probability $\frac{\psi_0}{\psi_0 + i}$ the new draw $\boldsymbol{\theta}_{i+1}$ is generated from the prior G_0, but it is otherwise equal to one of the previous i draws. When $\psi_0 \to \infty$ we have the same parametric model for each forecaster: $\mathbf{z}_i \sim h(\cdot|\boldsymbol{\theta}_i)$, where the $\boldsymbol{\theta}_i$'s are drawn independently from G_0.

$h(\mathbf{z}|\boldsymbol{\theta}_k^*)$. The Bayesian nonparametric model is therefore quite flexible. As such, it can overcome the inherent misspecification implied by the use of specific parametric assumptions, as shown below in section II.E. At the same time, the stick-breaking prior (10) on the mixture weights w_k imposes some degree of pooling, which mitigates overfitting—most forecasters come from the same relatively few types, with the number of types growing as the number of forecasters in the sample increases (a priori the expected number of clusters is approximately $\psi_0 \log((\psi_0 + n)/\psi_0)$). *A posteriori*, the number of mixtures used depends on the degree of heterogeneity in the sample, and both the unknown types $\boldsymbol{\theta}_k^*$ and the weights w_k are estimated, as described next.

II.D Posterior inference

In order to describe the posterior distribution, it is useful to rewrite the prior (11) using auxiliary allocation variables d's, which are equal to k if $\boldsymbol{\theta}_i$ is sampled from the k^{th} mixture component:

$$\mathbf{z}_i|G \overset{iid}{\sim} \sum_{k=1}^{\infty} \mathbb{I}\{d_i = k\} h(\mathbf{z}_i|\boldsymbol{\theta}_k^*), \quad Pr\{d_i = k\} = w_k. \tag{12}$$

Thus, the posterior distribution of $(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_n, G)$ given $(\mathbf{z}_1, ..., \mathbf{z}_n)$ can be expressed in terms of the posterior distribution $\Pi(d_1, ..., d_n, \boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, ..., w_1, w_2, ...|\mathbf{z}_1, ..., \mathbf{z}_n)$. If the mixture (12) were finite, Bayesian inference would be straightforward. The slice Gibbs sampler algorithm of Walker (2007) and Kalli et al. (2011) surmounts the issue of infinity using data augmentation, as explained in detail in Appendix B. The Markov Chain Monte Carlo samples

$$(d_1^{(m)}, ..., d_n^{(m)}, \boldsymbol{\theta}_1^{*(m)}, \boldsymbol{\theta}_2^{*(m)}, ..., w_1^{(m)}, w_2^{(m)}, ...)$$

over $m = 1, ..., M$ iterations are used to approximate the posterior distribution for any quantity of interest. For example, the set of posterior draws $\{F(y|\boldsymbol{\theta}_i^{(m)}) : y \in \mathcal{Y}, m = 1, ..., M\}$, with $\boldsymbol{\theta}_i^{(m)} := \boldsymbol{\theta}_{d_i^{(m)}}^{*(m)}$ approximates the posterior distribution of the subjective CDF $F_i(\cdot)$ (see Figure 2 below). Analogously, the posterior mean of the standard deviation of the predictive distribution of the i-th forecaster is approximated by

$$\frac{1}{M} \sum_{m=1}^{M} \sigma(\boldsymbol{\theta}_i^{(m)}),$$

where $\sigma(\boldsymbol{\theta})$ is the standard deviation of $F(\cdot|\boldsymbol{\theta})$. Finally, quantities involving the whole population of forecasters can be approximated in a similar way. For example, the posterior

mean of the cross-sectional standard deviation of the individual standard deviations is given by

$$\frac{1}{M}\sum_{m=1}^{M}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\sigma(\boldsymbol{\theta}_i^{(m)})-\frac{1}{n}\sum_{i=1}^{n}\sigma(\boldsymbol{\theta}_i^{(m)})\right)^2\right)^{1/2}.$$

II.E Posterior consistency

In this section we discuss asymptotic properties of the posterior distribution as the number of forecasters goes to infinity. We only state the main result on consistency, leaving all the details, proofs, and various additional results to Section C of the appendix.

We formalize asymptotic convergence using the notion of weak consistency of the posterior distribution (Ghosh and Ramamoorthi, 2003), which provides a widely accepted minimal requirement for large sample behavior of Bayesian nonparametric models (e.g., Norets and Pelenis, 2012; Pelenis, 2014; Norets and Pelenis, 2014; Bassetti et al., 2018). Roughly speaking, posterior consistency means that in a frequentist experiment with a given data-generating density, the posterior distribution concentrates around this density as the sample size (number of forecasters) increases. More formally, let $\mathcal{H}$ be the set of all possible data-generating densities (with respect to a dominating measure) on the simplex Δ^J where the data $\mathbf{z}$ lives. Given a prior Π on $\mathcal{H}$, the posterior $\Pi(\cdot|\mathbf{z}_1,...,\mathbf{z}_n)$ is said to be *weakly consistent* at a true density h_0 if for every i.i.d. sequence $\mathbf{z}_1,...,\mathbf{z}_n$ of random vectors with common density h_0 the posterior probability $\Pi(U|\mathbf{z}_1,...,\mathbf{z}_n)$ converges a.s. to 1 as $n\to+\infty$ for every weak neighborhood U of h_0. In our model, the prior is $\Pi(U)=P\{h_G\in U\}$, where $h_G(\mathbf{z})=\int h(\mathbf{z}|\boldsymbol{\theta})G(d\boldsymbol{\theta})$, see (11). In order to prove weak consistency we use the Schwartz theorem (see, e.g., Chapter 4 in Ghosh and Ramamoorthi, 2003), which states that weak consistency at a true density h_0 holds if the prior assigns positive probabilities to Kullback-Leibler neighborhoods of h_0.

Before stating the main theorem, we need to clarify the definition of densities and Kullback-Leibler divergence over a simplex Δ^J with possible zero elements. Recall that $\mathbf{z}_{\boldsymbol{\xi}}$ is the collection of strictly positive elements of $\mathbf{z}$ and that $\boldsymbol{\xi}$ is a vector indexing these positive z_j's. The Lebesgue measure on the sub-set of Δ^J identified by $\boldsymbol{\xi}$ is denoted by $\mathcal{L}_{\boldsymbol{\xi}}$ and a σ-finite measure on Δ^J can be defined as $\lambda(d\mathbf{z})=\sum_{\boldsymbol{\xi}}\mathcal{L}_{\boldsymbol{\xi}}(d\mathbf{z}_{\boldsymbol{\xi}})$. The set of all possible data-generating densities $\mathcal{H}$ is the set of all the densities $g(\mathbf{z})=g(\mathbf{z}_{\boldsymbol{\xi}},\boldsymbol{\xi})$ absolutely continuous with respect to λ. Given two densities h_0 and g in $\mathcal{H}$, the Kullback-Leibler divergence

between h_0 and g is then defined as

$$KL(h_0, g) = \int_{\Delta^J} h_0(\mathbf{z}) \log\Big(\frac{h_0(\mathbf{z})}{g(\mathbf{z})}\Big)\lambda(d\mathbf{z}). \tag{13}$$

Call $\mathcal{M}^*$ the set of finite (but arbitrarily large) mixtures of densities (5) that define the parametric component of our model, and $\mathcal{H}_0^*$ the set of densities that can be approximated in the Kullback-Leibler sense by densities in $\mathcal{M}^*$, i.e.,

$$\mathcal{H}_0^* = \{h_0 \text{ density w.r.t. } \lambda\text{: } \forall\, \epsilon > 0\ \exists\, g \in \mathcal{M}^* \text{ s.t. } KL(h_0, g) \leq \varepsilon\ \}.$$

Theorem 1. Assume that $\boldsymbol{\theta} \mapsto (\varrho_1(\boldsymbol{\theta}), ..., \varrho_J(\boldsymbol{\theta}), \phi\nu_1(\boldsymbol{\theta}), ..., \phi\nu_J(\boldsymbol{\theta}))$ is a continuous function such that $\nu_j(\boldsymbol{\theta}) > 0$ and $0 < \varrho_j(\boldsymbol{\theta}) < 1$ for every $j = 1, ..., J$ (recall that $\varrho_j(\boldsymbol{\theta})$ is the probability of reporting a zero for bin j). If G_0 has full support, then the posterior is weakly consistent at any density h_0 in $\mathcal{H}_0^*$ under suitable moment conditions.

The result of this theorem, which is stated more formally and proven in Appendix C, guarantees that the posterior distribution concentrates around the true process generating the histogram data $\mathbf{z}$ in the SPF cross-section as the number of forecasters grows to infinity. In particular, it shows that the Bayesian nonparametric approach is robust to deviations from the specific parametric assumptions, such as the particular choice of the $F(\cdot)$ predictive CDF. Hence, even if the specific form of $h(\mathbf{z}|\boldsymbol{\theta})$ is not correct, the true distribution h_0 is recovered in the limit as long as h_0 belongs to the very broad class of models $\mathcal{H}_0^*$, which includes all the models that are not "too far" from any finite mixture of $h(\mathbf{z}|\boldsymbol{\theta})$. We claim that this property is not shared by any of the current outstanding approaches for inference on probabilistic surveys.[10]

A couple of observations are in order. First, as is always the case for Bayesian nonparametrics, the consistency results do not apply to *individual* forecasters, but only to the data-generating process for the entire distribution of forecasters. Concretely, this means

[10] How large is the set $\mathcal{H}_0^*$ the true h_0 must belong to in order for Theorem 1 to have bite? It is a known result that a mixture of Dirichlet distributions can approximate any continuous distributions on the simplex. Proposition 2 in Appendix C shows that, as long as the $F(\cdot|\theta)$ distribution is flexible enough (e.g., is a large enough mixture of Gaussians), this results carries over to our setting. In our applications, $F(\cdot|\theta)$ consists of a mixture of two normals, so we do impose some parametric restrictions that make $\mathcal{H}_0^*$ less general. In practice though, i) we think that a mixture of two normals is broad enough to cover most realistic cases, and ii) we consider robustness to using three normals and show that this does not meaningfully change the results. Proposition 2 considers the case without rounding toward zero. Adding distributions that place discrete mass on zeros makes the class $\mathcal{H}_0^*$ larger.

that they apply to any object that involves a suitably large number of forecasters, such as the consensus distribution. Second, it must be clear that the consistency holds for the true distribution h_0 on the available data $\mathbf{z}$ and *not* for the underlying predictive CDF $F(\cdot)$ over the entire domain of y. This is due to the fact that the available data do not provide enough information to fully recover the CDF of y since the number of bins J is taken as fixed (and finite), even when n goes to infinity: loosely speaking, we can claim consistency for the value of the predictive CDF $F(\cdot)$ at the bin edges $y_1, ..., y_J$, but we do not have enough information about the value of $F(\cdot)$ for $y \in (y_j, y_{j+1}]$. This identification issue is overcome in the case where the number of bins J goes to infinity and the bin size goes to zero, as we show in Appendix C. Specifically, we show that, under these conditions, when the number of forecasters n also goes to infinity the consistency result discussed above applies also to estimates of the predictive distribution $F(\cdot)$. Since these results are of limited interest for our application where the number of bins is limited and non-negligible mass is often placed on the open bins, they are relegated to the appendix.[11]

In practice, in our application both the number of bins and of forecasters n are not large (e.g., n is around 30 in 2020 and the width of some of the bins is as large as 6 percent for output growth). Still, an advantage of the Bayesian approach is that lack of information is reflected in the posterior credible intervals. Section III.A below, for instance, shows that when forecasters place a large amount of mass on the open bins, the estimates of their predictive CDF $F(\cdot|\boldsymbol{\theta}_i)$ become more uncertain. Still, one needs to be aware that in these situations the choice of the distribution family $F(\cdot)$, $h(\cdot)$, and the prior distribution can impact the results. Therefore, a robustness check with respect to these choices should be included in all applications of our method and we do so in this paper, as discussed below.

II.F Priors

In this section we discuss the prior settings used in our application. Recall that $\boldsymbol{\theta} = (\mu, \mu_\delta, \sigma_1, \sigma_2, \omega, \phi, \epsilon)$. The first four parameters pertain to the $F(\cdot)$ function—the mixture of two normals (A-26), which we repeat here for convenience: $F(y|\boldsymbol{\theta}) = (1-\omega)\Phi(y|\mu, \sigma_1^2) + \omega\Phi(y|\mu + \mu_\delta, \sigma_2^2)$. The parameters ϕ and ϵ are used to specify $h(\cdot)$ and $\varrho(\cdot)$ in (6) and (7),

[11] Arguably, SPF surveys of output growth between the Great Recession and the Covid episode, which displayed fairly narrow bins and little mass on the open bins, are the only ones for which these conditions come close to applying.

respectively. We should stress that we use the same priors for both output growth and inflation, and for all years in our sample.

The location of the first mixture component is $\mu \sim \mathcal{N}(2, 5^2)$, where the standard deviation of 5 implies that this is a very loose prior. The scales of the mixture components follow $\sigma_j \sim \mathcal{IG}a(a_\sigma, b_\sigma)\mathbb{I}_{(0,10)}(\sigma_1)$, $j = 1, 2$, where a_σ, b_σ are chosen so that the standard deviation has a mean $\mathbb{E}[\sigma_j] = 2$ and a variance $\mathbb{V}[\sigma_j] = 4$, and where we truncate the distribution at 10 for numerical reasons. The parameter μ_δ captures the deviation of the mean of the second mixture component relative to the first one. Its prior is centered at zero (implying that the second mixture a priori mainly captures fat tails) and has a standard deviation of 1: $\mu_\delta \sim \mathcal{N}(0, 1^2)$. The prior for ω, the weight on the second component of the mixture, is $\omega \sim \mathcal{B}e(0.5, 3)$. Its mode is zero, implying that the prior favors models with one mixture only. The prior places roughly 20 percent probability on $\{\omega \geq 0.25\}$.

For the precision parameter ϕ of the Dirichlet distribution (6)—recall that ϕ determines the amount of noise around the underlying probabilities ν—we assume $\phi \sim \mathcal{G}a(a_\phi, b_\phi)\mathbb{I}_{(\underline{\phi},\infty)}(\phi)$, where a_ϕ and b_ϕ are chosen so that at the prior mean, when the underlying probability ν is 50 percent, the noise is 2.5 percent. $\underline{\phi}$ is chosen so that the noise has an upper bound of 5 percent. The left panel of Figure A-2 in the appendix shows the 50 and 90 percent a-priori coverage intervals for the noise associated with three different values of ν: 0.1, 0.6, and 0.3. The 50 and 90 percent intervals are about 5 and 10 percent wide, respectively. Regarding the prior for the rounding-off-to-zero parameter ϵ, we assume a truncated gamma distribution $\mathcal{G}a(a_\epsilon, b_\epsilon)\mathbb{I}_{(0,\bar{\epsilon})}(\epsilon)$ and set a_ϵ, b_ϵ such that at the prior mean the probability of reporting a zero, ϱ, is 5 percent when the underlying ν is 2.5 percent. We choose $\bar{\epsilon}$ so that ϱ is at most 20 percent for ν equal to 2.5 percent. The right panel of Figure A-2 shows the mean, the 50 percent, and the 90 percent coverage intervals of ϱ as a function of ν. The a-priori uncertainty is such that when ν is 2 percent, the 90 percent interval for ϱ ranges from 0 to 25 percent, with the coverage interval converging to zero for ν greater than 4 percent. Finally, for the concentration hyperparameter ψ_0 of the Bayesian nonparametric prior, which determines the prior number of clusters, we follow the standard choice and set $\psi_0 = 1$. This implies that the expected number of clusters for a cross-section with $n = 30$ is roughly 4.

III Results

In this section we first discuss the application of the Bayesian nonparametric approach to the few selected examples mentioned at the beginning of Section II, in order to familiarize the reader with how the approach works in practice. Next, we document the evolution from 1982 to 2022 of individual measures of uncertainty obtained using our approach. This analysis sets the stage for the analysis in the following section, where we study the relationship between subjective uncertainty and ex-post forecast errors, and assess whether SPF predictive densities conform with the noisy RE hypothesis.

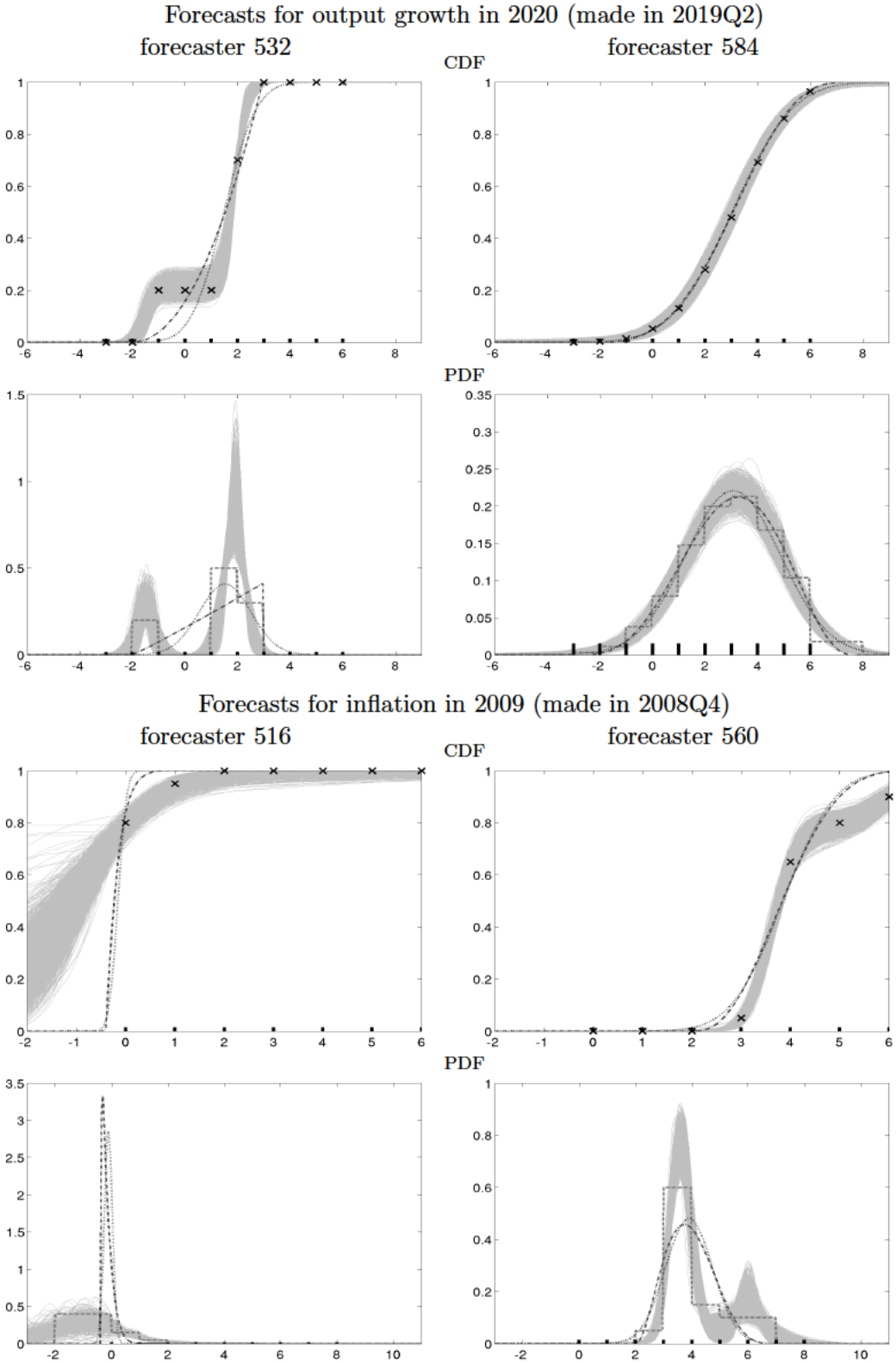
III.A Examples

Figure 2 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 Bayesian nonparametric model for the subjective CDF $F(y|\boldsymbol{\theta}_i)$ (top) and PDF (bottom), and we compare the results with those obtained under the generalized beta (black dash-and-dotted lines) and Gaussian (black dotted lines) approaches. The CDF plots also show the observed cumulative probabilities Z_{ij} (crosses), while the PDF plots show the step-wise uniform PDF (gray dashed lines) implied by the histogram probabilities z_{ij}.

Figure 2 illustrates a few aspects of the Bayesian nonparametric approach. First, the observed cumulative probabilities (the Z_{ij}'s; crosses) belong to the high-posterior-density region for all these respondents, suggesting that the approach is flexible enough to capture a variety of arguably challenging cases. In contrast, the beta and the normal approaches do not fit the Z_{ij}'s well in most of these examples, and their CDFs and PDFs do not belong to the high-posterior-density region obtained from the Bayesian nonparametric approach, with the exception of respondent 584. As a consequence, there can be important differences in the objects of interest, such as the measure of uncertainty, or quantiles, implied by the different approaches.[12] Figure 2 also shows that the Bayesian nonparametric approach delivers wider posterior coverage intervals that reflect the higher degree of uncertainty whenever there is less information from the respondent. The case of respondent 516, who placed 80 percent probability on the left open bin (see Figure 1), is exemplary. The posterior coverage intervals for both the Bayesian nonparametric CDF and PDF reflect the fact that we know very little

[12] Bassetti et al. (2023) show that inference using the Bayesian nonparametric approach differs from that obtained using standard approaches for several other examples obtained during the recent Covid episode.

Figure 2: Inference using the Bayesian nonparametric approach: CDFs and PDFs for selected examples



Note: For each forecaster the top and bottom panels show the subjective CDF and PDF, respectively, estimated using the Bayesian nonparametric approach (posterior draws; light gray), as well as the least-squares estimates obtained under the normal (gray, dashed line) and the beta (black, dash-and-dotted line) parametric assumptions. The CDF panels also show the observed cumulated histogram probabilities Z_{ij} $j = 1, \ldots, J$ (crosses), while the PDF panel shows the step-wise uniform PDF (gray dashed lines) implied by such probabilities.

about the left-tail behavior of this forecaster, 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 in other regions of the distribution or for other forecasters.

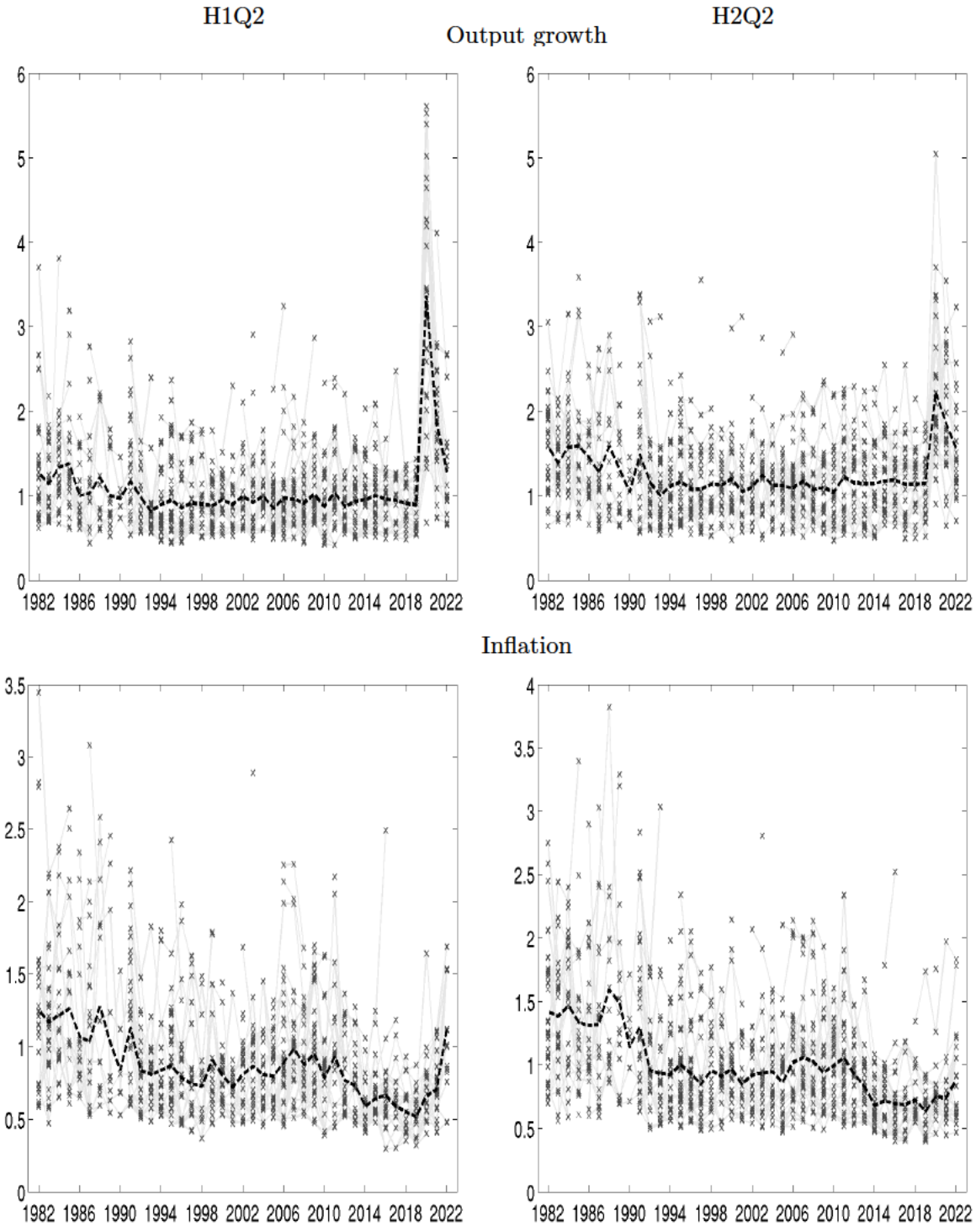
III.B Heterogeneity in subjective uncertainty

In this section we document the evolution of individual measures of uncertainty obtained using our approach in the 1982-2022 sample. We do this for two reasons. First, we set the stage for the analysis in the next section, where we study the relationship between subjective uncertainty and ex-post forecast errors. In particular, we show that professional forecasters differ significantly in terms of their assessment of uncertainty, and that these differences vary over time.[13] We also show that while these differences are persistent, forecasters do change their minds from period to period about their subjective uncertainty—a variation that we will exploit later. Second, we take advantage of our inference-based approach and test the extent to which these differences are significant.

Figure 3 shows the evolution of subjective uncertainty by individual respondent for output growth (top) and inflation (bottom). The left and right panels display uncertainty for the current and the next-year projections, respectively, made in the second quarter of each year (Figures A-8 through A-10 in the appendix show that results for other quarters are similar in terms of the features discussed in this section). In each panel, the crosses indicate the posterior mean of the standard deviation of the individual predictive distribution. We use the standard deviation (as opposed to the variance) because its units are easily grasped quantitatively and are comparable with alternative measures of uncertainty such as the interquartile range. Thin gray lines connect the crosses across periods when the respondent is the same, providing information as to whether respondents change their view on uncertainty and whether the composition of the panel affects the cross-sectional average measure of

[13] Heterogeneity in macroeconomic probabilistic forecasts was noted long ago. While much of the early literature focused on disagreement in point projections or central tendencies (see Mankiw et al., 2003; Capistrán and Timmermann, 2009; Patton and Timmermann, 2010, 2011; Andrade and Le Bihan, 2013; and other work mentioned in the recent survey by Clements et al., 2023), more recent work documents the fact that forecasters disagree about uncertainty and that these differences are long-lasting (Lahiri and Liu, 2006; D'Amico and Orphanides, 2008; Bruine De Bruin et al., 2011; Boero et al., 2014; Rich and Tracy, 2021, among others). Kozeniauskas et al., 2018, discuss the conceptual differences between macroeconomic uncertainty and disagreement using a model where forecasters have private information and update their beliefs using Bayes law.

Figure 3: Subjective uncertainty by individual respondent



Note: Each panel displays the posterior mean of the standard deviation of the subjective predictive distribution by individual respondent (light gray crosses, connected by thin gray line whenever the respondent appears in consecutive surveys), and the cross-sectional average of the individual standard deviations (dashed black line). Top and bottom panels correspond to output growth and inflation projections. The left and right columns correspond to current and next-year projections.

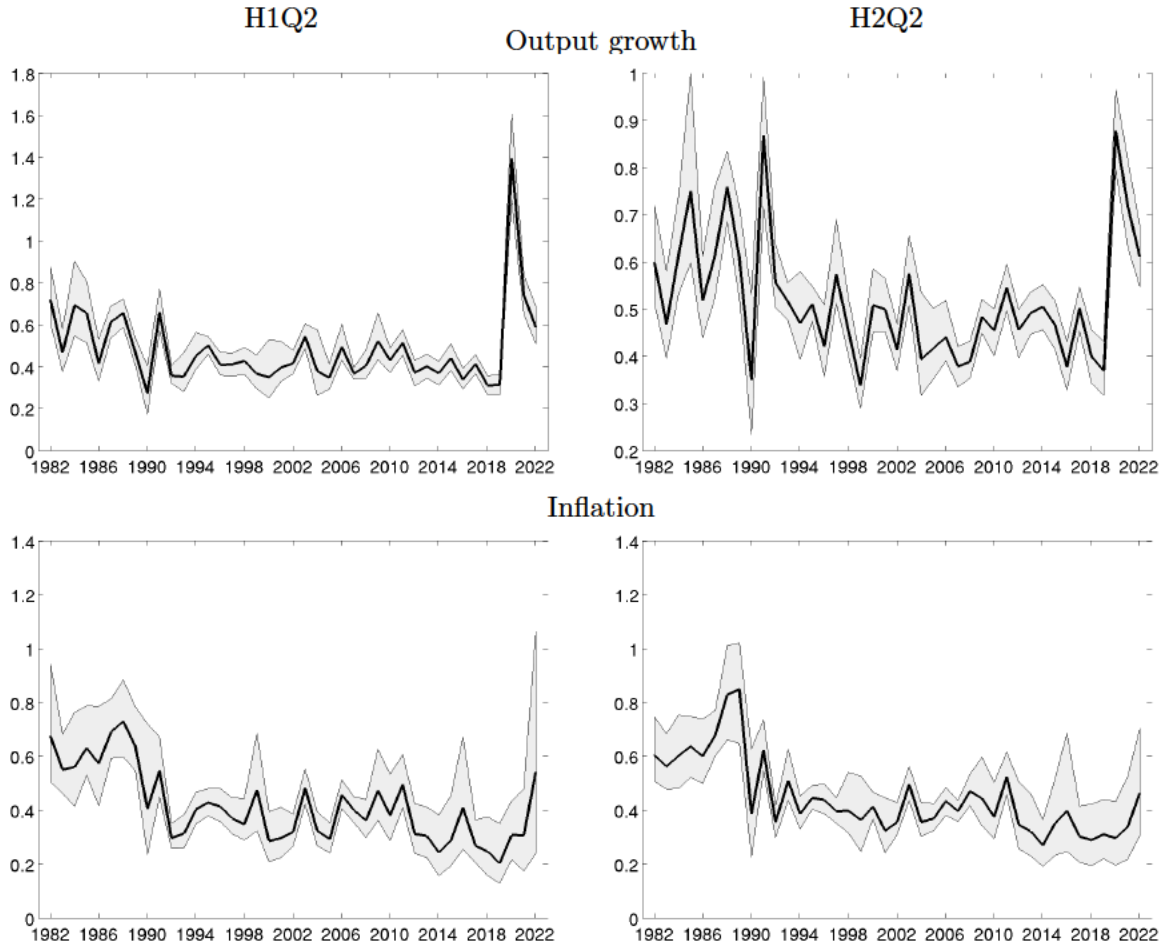
uncertainty, which is shown by a black dashed line (Manski, 2018, stresses the extent to which the literature has often ignored compositional changes when discussing the evolution of consensus or average measures). Figure 4 provides a time series of the differences in individual uncertainty, as measured by the cross-sectional standard deviation of the individual standard deviations. The solid black line displays the posterior mean of this measure, while the shaded areas represent the 90 percent posterior coverage.

Figure 3 shows that, on average, uncertainty for current-year output growth projections declined from the 1980s to the early 1990s, likely reflecting a gradual learning about the Great Moderation, and then remained fairly constant up to 2020, when the Covid pandemic struck and average uncertainty grew threefold. Average uncertainty for next-year projections tends to be higher than for current-year projections. It follows a similar pattern, except that it displays a small but very steady upward shift in the aftermath of the Great Recession. It appears unlikely that changes in survey design, and particularly in the bins, affect these patterns: for output growth, these changes take place in 1992, 2009, and 2020. Except for 2020, where much of the change in uncertainty is arguably attributed to Covid, there are no evident breaks associated with the bin changes. Interestingly, we do not see any upticks in average subjective uncertainty in the run-up to recessions, even for current-year forecasts, with the exception of the Covid crisis. We obtain very similar results if we use the interquantile range to measure uncertainty (Figure A-11). Using the generalized beta approach to fit histograms (Figure A-12) also produces similar overall patterns, although this approach leads to lower estimates of subjective uncertainty relative to our approach.

Cross-sectional differences in individual uncertainty are very large and quantitatively trump any time variation in average uncertainty. The standard deviation of low-uncertainty individuals remains below one throughout the sample, with the sole exception of the Covid period, while that of high uncertainty individuals is often higher than two. More formally, the cross-sectional standard deviation of individual standard deviations, shown in Figure 4, hovers between 0.4 and 0.8 throughout the sample and then jumps during the Covid period. The cross-sectional standard deviation is quite tightly estimated, indicating that differences across individuals are significant. The level and the dispersion of uncertainty appear to be linked, in that the cross-sectional standard deviation of uncertainty is high when the average uncertainty is high. From Figure 3, this result seems due to the fact that it is mostly high-uncertainty respondents who change their minds about the confidence in their projections, thereby driving both the average and the cross-sectional standard deviations. While differences in subjective uncertainty are persistent, forecasters do revise their assessment of

unpredictability, sometimes substantially (see the thin gray lines in Figure 3).

Figure 4: Cross-sectional standard deviation of individual uncertainty—Q2 survey



Note: Posterior mean of the cross-sectional standard deviation of the individual standard deviations (solid black line) The shaded areas display the 90 percent posterior credible intervals. Top panels: output growth. Bottom panels: inflation. Left column: current year; right column: next year.

The bottom panels of Figure 3 show that, on average, subjective uncertainty for inflation in both the current (left) and following (right) years declined from the 1980s to the mid-1990s and then was roughly flat up until the mid-2000s. Average uncertainty rose in the years surrounding the Great Recession, but then declined again quite steadily starting in 2011, reaching a lower plateau around 2015. Interestingly, average uncertainty did not rise dramatically in 2020 through 2022 despite Covid and its aftermath, and despite the fact that, for most respondents, expected inflation rose sharply (Figure A-14). In the case of inflation, changes in the bins, which took place in 1985, 1992, and 2014 (Figure A-3), may have played some role, as we see that the average standard deviation drops markedly in both 1992 and 2014. At the same time, it is arguably not the only explanation since such drops appear to be the continuation of a trend that started before the change in survey design.

For inflation, cross-sectional differences in individual uncertainty are also very large. The cross-sectional standard deviation of individual standard deviations (Figure 4) follows the same pattern of the average standard deviation: it starts around 0.6 percent in the 1980s, drops to around 0.4 percent in the 1990s, and then drops a bit further in the 2010s. This measure of cross-sectional heterogeneity in uncertainty is tightly estimated and its fluctuations are statistically significant. As was the case for output, high-uncertainty respondents becoming less uncertain are mostly driving both the average and the cross-sectional standard deviations.

III.C Subjective uncertainty and forecast accuracy: Testing the noisy rational expectations hypothesis for density forecasts

In this section we use our approach to assess whether SPF predictive densities conform with the noisy RE hypothesis (see Coibion and Gorodnichenko, 2012, 2015, among many other references). According to this hypothesis, forecasters receive both public and private signals about the state of the economy, which they do not observe. Heterogeneity in the signals, and in their precision, explains the heterogeneity in both mean predictions $\mathbb{E}_{t-q,i}[y_t]$ and in their subjective uncertainty $\sigma^2_{t|t-q,i} = \mathbb{E}_{t-q,i}[(y_t - \mathbb{E}_{t-q,i}[y_t])^2]$, where i denotes the forecaster, q the horizon of the forecast, and $\mathbb{E}_{t-q,i}[\cdot]$ is the expectation operator under forecaster i's information set at time $t-q$. In the time series, changes in the precision of either the private or public signals—the latter due, say, to changes in policy or the structure of the economy, a recession approaching, or some other major event like Covid-19—will be reflected in $\sigma^2_{t|t-q,i}$. In the cross-section, if forecaster i has a more precise signal than forecaster j, then $\sigma^2_{t|t-q,i}$ ought to be lower than $\sigma^2_{t|t-q,j}$. We have seen in the previous section that $\sigma^2_{t|t-q,i}$ varies substantially both over time and in the cross-section.

We assess the noisy RE hypothesis using three types of tests—the first two concerning the scale of the forecasters' predictive distribution, and the last concerning its location. The first two tests are based on the idea that if expectations are rational there needs to be a correspondence between the subjective uncertainty $\sigma_{t|t-q,i}$ and the *ex-post* forecast error $|y_t - \mathbb{E}_{t-q,i}[y_t]|$. Define

$$\eta_{i,t|t-q} = (y_t - \mathbb{E}_{t-q,i}[y_t])/\sigma_{t|t-q,i}, \tag{14}$$

the standardized forecast error. Under RE (that is, if the predictive distribution is consistent with the data-generating process for y_t) it has to be the case that

$$\mathbb{E}[\eta^2_{i,t|t-q}] = 1. \tag{15}$$

We will test whether $\eta^2_{i,t|t-q} = (y_t - \mathbb{E}_{t-q,i}[y_t])^2/\sigma^2_{t|t-q,i}$, is equal to 1 on average, and refer to this test as a *scale* test. Next, taking logs of the absolute value of both sides of equation (14) we obtain

$$\ln|y_t - \mathbb{E}_{t-q,i}[y_t]| = \ln\sigma_{t|t-q,i} + \ln|\eta_{i,t|t-q}|. \tag{16}$$

By regressing $\ln|y_t - \mathbb{E}_{t-q,i}[y_t]|$ on $\ln\sigma_{t|t-q,i}$ we will test whether absolute forecast error changes proportionally to the subjective uncertainty both in the time series and in the cross-section. We refer to this test as a *variation* test. While both tests hinge on RE, they are different. The *scale* test tells us whether subjective distributions are properly scaled on average, while the *variation* test inquires whether variations in subjective uncertainty map into variations in forecast errors. Aside from testing the noisy RE hypothesis, this latter test is interesting in itself, as it sheds light on the relationship between the *ex-ante* uncertainty expressed by survey respondents and their *ex-post* ability to predict macroeconomic outcomes, both in the time series and in the cross-section.

Finally, under RE it has to be the case that the mean projection $\mathbb{E}_{t-q,i}[y_t]$ leads to smaller forecast errors on average than any other forecast. This is because the mean projection minimizes the expected squared forecast error under $\mathbb{E}_{t-q,i}[\cdot]$ and therefore, under RE, under the unconditional expectation $\mathbb{E}[\cdot]$ as well:

$$\mathbb{E}[(y_t - \mathbb{E}_{t-q,i}[y_t])^2] \leq \mathbb{E}[(y_t - y^{pp}_{t,t-q,i})^2] \text{ for any } y^{pp}_{t,t-q,i}. \tag{17}$$

We refer to this test as a *location* test, as it assesses whether the predictive densities' mean fulfills its properties under RE. A strand of literature has investigated whether point forecasts coincide with means and, to the extent that they do not, whether this reflects a forecaster's loss function that is not quadratic (e.g., Engelberg et al., 2009; Clements, 2010; Elliott et al., 2008; Patton and Timmermann, 2007). Our test is based on the notion that regardless of the forecaster's loss function, the mean better minimize the square loss if forecasters are rational.[14]

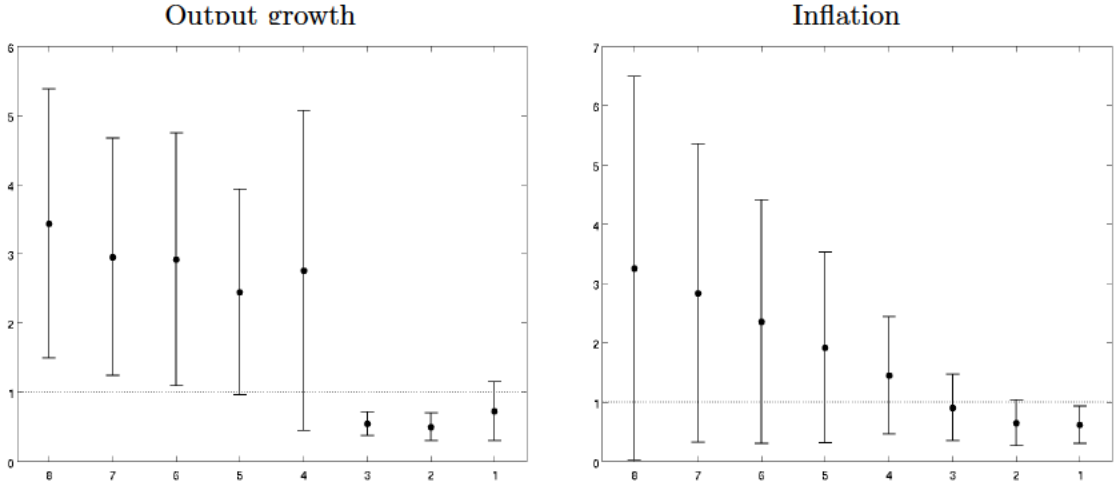
III.C.1 A scale test: Do forecasters over or underestimate uncertainty?

We can assess the hypothesis in (15) by testing whether $\alpha_q = 1$ in the panel regression

$$(y_t - \mathbb{E}_{t-q,i}[y_t])^2/\sigma^2_{t|t-q,i} = \alpha_q + \epsilon_{t,i,q},\ t = 1, ..., T,\ i = 1, ..., n, \tag{18}$$

[14] Another strand of the literature (e.g., Ganics et al., 2023, and references therein) has used probabilistic SPF predictions to guide model-based density forecasts, but this literature almost exclusively uses the consensus SPF predictions.

Figure 5: Do forecasters over- or underestimate uncertainty? A scale test



Note: Black dots correspond to OLS estimates of α_q from regression (18) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors. Table A-1 in the appendix provides the regression statistics.

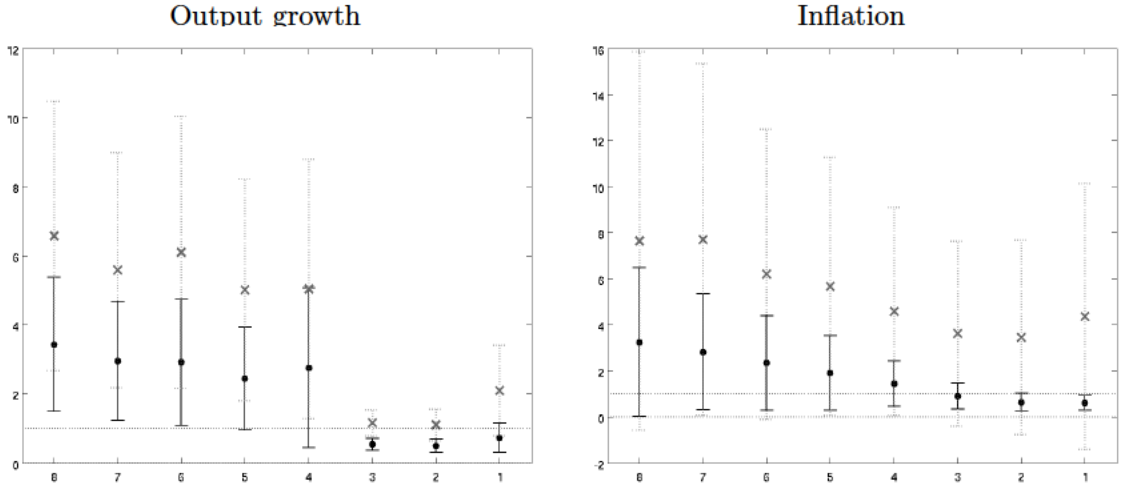
where we use the posterior means of $\mathbb{E}_{t-q,i}[y_t]$ and $\sigma^2_{t|t-q,i}$ from our approach. Estimates of α_q that are significantly greater/lower than one indicate that forecasters under/over-estimate uncertainty. Figure 5 shows estimates of α_q for different horizons, ranging from $q = 8$ (H2Q1) to $q = 1$ (H1Q4). (Recall that the variables being forecast are the year-over-year growth rates of output or the price level.)[15] The dots indicate the OLS point estimates and the whiskers indicate the two-standard-deviations posterior intervals, computed using From Fixed-Event to Fixed-Horizon Density Forecasts: Obtaining Measures of Multihorizon Uncertainty from Survey Density ForecastsDriscoll and Kraay (1998) standard errors, which are robust to heteroskedasticity and serial and cross-sectional dependence including clustering by time and forecaster.

The results are as follows. For long horizons—between two and one years—forecasters are overconfident by a factor ranging from two to four for both output growth and inflation. The opposite is true for short horizons: on average forecasters overestimate uncertainty, with point estimates of α_q lower than one for q less than four quarters. The standard errors are large, especially for long horizons. For output growth the estimates are significantly above one for q greater than six, but for inflation the 90 percent coverage intervals always include

[15] In line with most of the literature evaluating forecasting accuracy from surveys, in this section we use the "first final" estimate as a measure for y_t, where the first final is the third estimate of GDP or the GDP deflator. Results obtained using the latest revision are shown in Figure A-18 of the appendix. These results are very similar for longer horizons but tend to be different for shorter horizons, arguably because the revised series contain methodological changes in measuring GDP that forecasters simply cannot anticipate.

one. Figure A-19 in the appendix shows that the pattern of overconfidence at long horizons and underconfidence at short horizons is robust across different sub-samples (excluding the Covid period and/or the period 1982-1991 when the Philadelphia Fed was not in charge of the survey), although the degree of overconfidence for long horizons changes with the sample, especially for inflation. While the findings are in line with the literature on overconfidence (Daniel and Hirshleifer, 2015; Malmendier and Taylor, 2015) for output for horizons greater than one year, results are more uncertain for inflation. For horizons shorter than three quarters, forecasters are underconfident for both variables.

Figure 6: A scale test—comparison with the generalized beta approach



Note: Black dots correspond to OLS estimates of α_q from regression (18) for $q = 8, ..., 1$ using the posterior means for $E_{t-q,i}[y_t]$ and $\sigma^2_{t|t-q,i}$ from our approach. Gray crosses correspond to OLS estimates when these objects are obtained using the generalized beta approach. Whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors. Table A-2 in the appendix provides the regression statistics.

The idea behind the regression in (18) builds on existing literature. Clements (2014) in particular computes values for $(y_t - \mathbb{E}_{t-q,i}[y_t])^2/\sigma^2_{t|t-q,i}$ using the point predictions in place of the mean $\mathbb{E}_{t-q,i}[y_t]$, and estimates of $\sigma^2_{t|t-q,i}$ obtained from fitting a generalized beta distribution. Clements then computes α_{iq} using a time-series regression for each forecaster i, tests the hypothesis $\alpha_{iq} > 1$ and $\alpha_{iq} < 1$, and reports the fraction of forecasters for which each hypothesis is rejected.[16] Broadly speaking, the findings of Clements (2014) are in line with those reported above: at longer horizons forecasters generally tend to be overconfident, and this overconfidence diminishes as the horizon gets shorter.

[16]This exercise is conducted for SPF surveys of output growth and inflation from 1981Q3 to 2010Q4. Giordani and Soderlind (2003) also find that forecasters underestimate inflation uncertainty. Similarly, Diebold et al. (1999) and Rossi and Sekhposyan (2015, 2019) investigate whether uncertainty is over or underestimated but focus on the consensus (that is, average) predictive density.

The benefit of running a panel regression as in (18) is twofold, compared to a forecaster-by-forecaster analysis. First, we obtain quantitative estimates of the average degree of over/underconfidence that are not marred by the small-sample problem affecting individual forecasters' regressions. Second, different from previous work this regression provides an explicit test of the RE hypothesis for SPF density forecasts, in line with the recent literature testing RE for point forecasts (e.g., Bordalo et al., 2020).[17] This test for instance reveals that the view that forecasters are overconfident about inflation is questionable, even for long horizons. Also, previous literature mostly used point forecasts, while of course under RE equation (15) holds for the mean, but not necessarily for the point forecast if this differs from the mean (Figure A-17 shows that the results for the point forecasts are not very different at long horizons, but can be quite different at short horizons). Finally, Figure 6 shows that it makes a big difference whether one uses the posterior mean of $\sigma^2_{t|t-q,i}$ from our approach or that obtained from fitting a generalized beta distribution, especially at long horizons where forecasters place more probability on the open bins.[18]

III.C.2 A variation test: Do differences in subjective uncertainty over time or across forecasters map into differences in forecast accuracy?

Next, we explore a different implication of the noisy RE hypothesis: subjective uncertainty and forecast accuracy should co-move, both across forecasters and over time (see equation (16)). We examine this hypothesis by testing whether in the panel regression

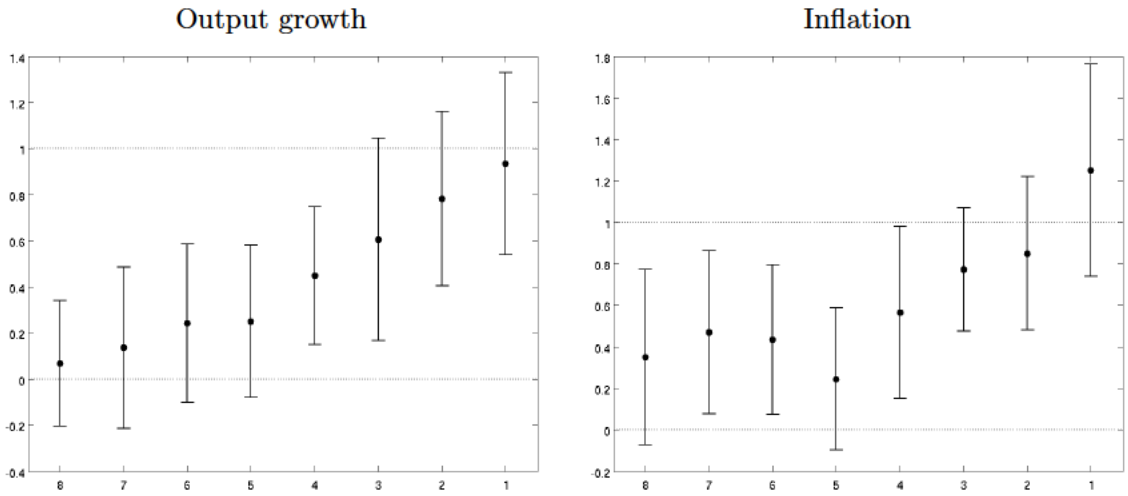
$$\ln|y_t - \mathbb{E}_{t-q,i}[y_t]| = \beta_{0,q} + \beta_{1,q} \ln \sigma_{t|t-q,i} + \epsilon_{t,i,q}, \;\; t = 1, ..., T, \;\; i = 1, ..., n, \tag{19}$$

the coefficient $\beta_{1,q}$ is equal to one. As before, equation (19) is estimated via OLS where $\mathbb{E}_{t-q,i}[y_t]$ and $\sigma_{t|t-q,i}$ are measured using the posterior mean of the standard deviation estimated using our approach, and Driscoll-Kraay standard errors are computed.

[17] The regressions in this section assume homogeneity in that they test for the *average* behavior of forecasters, again in line with the literature. An interesting research agenda, to which Clement's work has contributed, would be to test for, and uncover the sources of, heterogeneity among forecasters.

[18] Glas and Hartmann (2022) conduct Clement's exercise by distinguishing between rounders (respondents who round up to zero and/or to round numbers) and non-rounders. They find that rounders tend to underestimate uncertainty, especially for long horizons, while non-rounders do not. Our approach takes rounding explicitly into account. In fact, we find that the underestimation of uncertainty at long horizons is about half that implied by the generalized beta approach.

Figure 7: Do differences in subjective uncertainty map into differences in forecast accuracy? A variation test

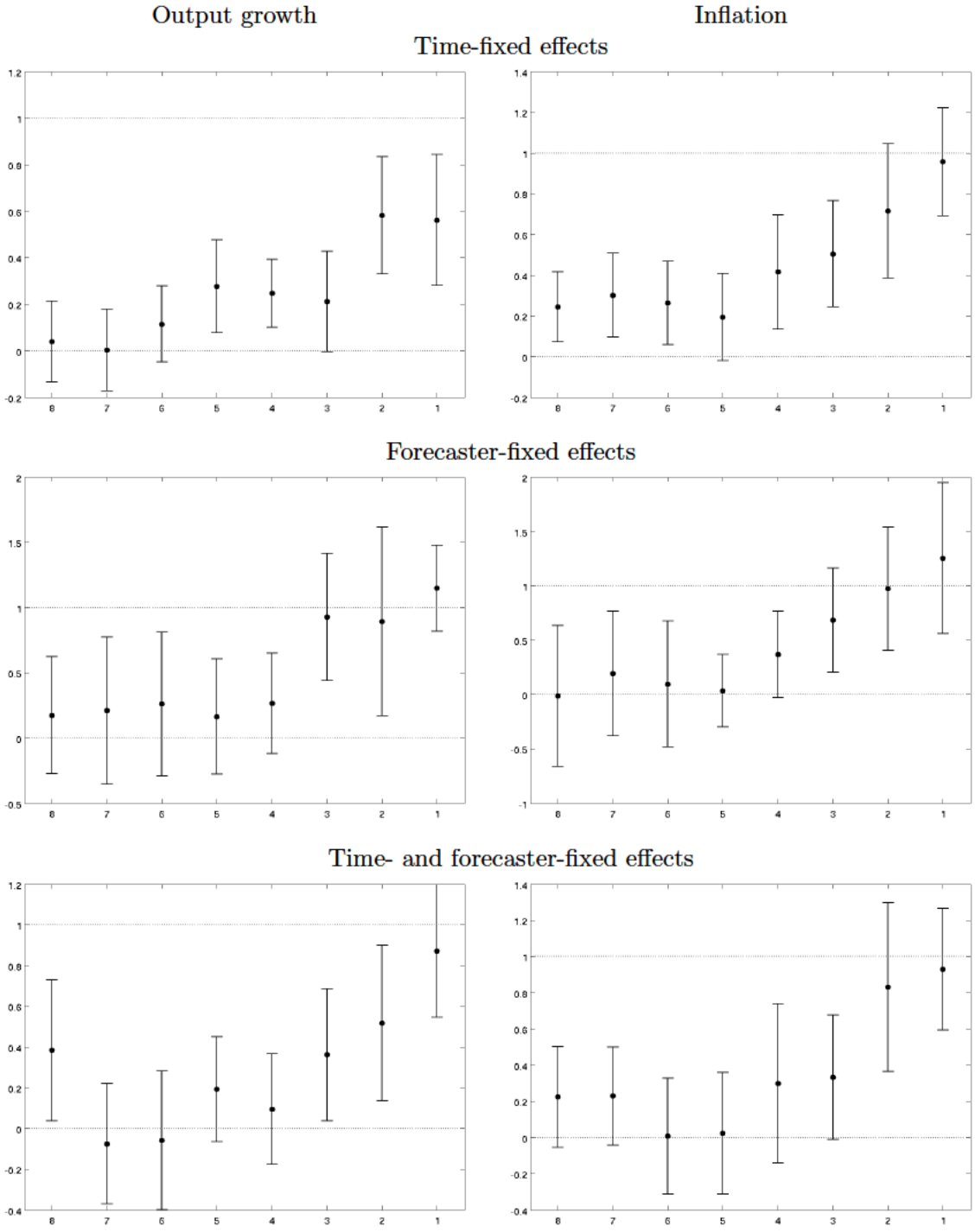


Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors. Table A-3 in the appendix provides the regression statistics.

Figure 7 plots the point estimates of $\beta_{1,q}$ for different horizons (crosses) and the whiskers denote the two-standard-deviations posterior intervals. For output growth we cannot reject the hypothesis that $\beta_{1,q} = 0$ for $q \geq 5$ and the point estimates are all less than 0.3: essentially there is no relationship between subjective uncertainty and the size of the *ex-post* forecast error for horizons beyond one year. As the forecast horizon shortens, the relationship becomes tighter, and for horizons of three quarters or less ($q \leq 3$) one cannot reject the hypothesis that $\beta_{1,q} = 1$, that is, one cannot reject RE. For inflation the point estimates of $\beta_{1,q}$ hover between 0.2 and 0.5 for longer horizons, but increase toward 1 as the horizon shortens, with estimates that are also not significantly different from 1 for horizons of three quarters or less. Figures A-24 through A-29 in the appendix show that the main finding—$\beta_{1,q}$ is well below one, and often indistinguishable from zero, for long horizons while one cannot reject $\beta_{1,q} = 1$ for short horizons, for both output and inflation—is robust to different samples and specifically does not depend on including the Covid period.

Figure 8 shows the estimates of $\beta_{1,q}$ controlling for time, forecaster, and both time- and forecaster-fixed effects in order to ascertain whether the results in Figure 7 are mostly due to differences across forecasters or over time. The results with time-fixed effects (top panels) indicate that for output growth forecasters with lower/higher subjective uncertainty do not have lower/higher absolute forecast errors on average at long horizons, although for short horizons the correspondence between the two becomes tighter. For inflation the relationship is also far from one-to-one at longer horizons. For $q \leq 5$ the coefficient $\beta_{1,q}$

Figure 8: A variation test—regressions with fixed effects

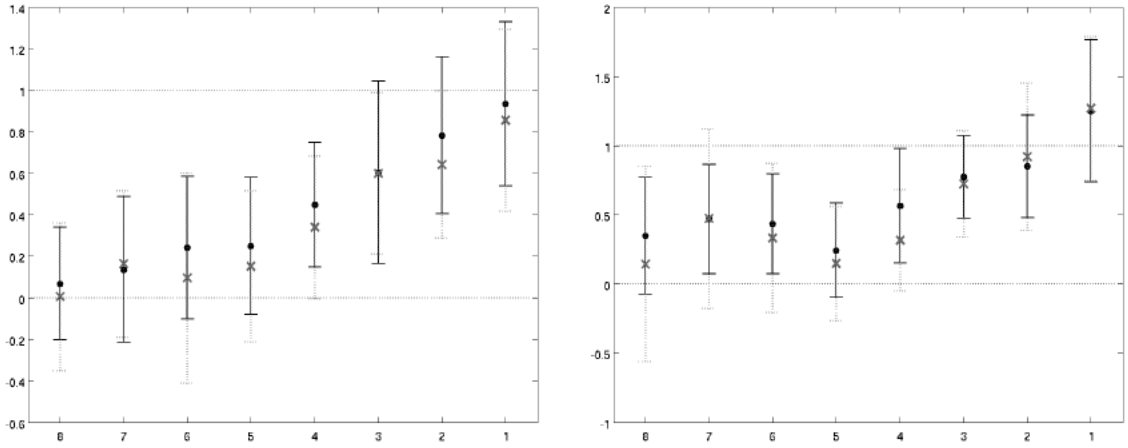


Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) using time-fixed effects (top panels), forecaster-fixed effects (middle panels), or both (bottom panels), for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors. Table A-3 in the appendix provides the regression statistics.

increases monotonically with q, and $\beta_{1,1}$ is indistinguishable from one.

The results with forecaster-fixed effects (middle panels) suggest that when forecasters

Figure 9: A variation test–accounting for inference uncertainty (baseline vs. weighted OLS)



Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors. Gray crosses correspond to weighted OLS estimates, where the weights are inversely proportional to inference uncertainty as measured by the interquantile range of the posterior distribution of $\sigma_{t|t-q,i}$. Whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors. Table A-5 in the appendix provides the regression statistics.

change their subjective uncertainty, because of aggregate uncertainty shocks (e.g., Covid) or because the quality of their private signal has changed, on average this maps one-to-one into corresponding changes in the absolute forecast errors for short horizons ($q \leq 3$), but has essentially no bearing for forecast accuracy for longer horizons. The bottom panels of Figure 8 show that these findings still hold when we include both forecaster- and time-fixed effects.

The idea of testing the noisy RE hypothesis using the variation test is novel. Clements (2014) computes time-series averages of $\sigma_{t|t-q,i}$ for each forecaster and plots them against the corresponding predictive root mean square error (RMSE) computed during the same period.[19] Clements concludes that "there is little evidence that more (less) confident forecasters are more (less) able forecasters." This exercise compares to our model with time-fixed effects, where we study whether forecasters that are more uncertain also have higher absolute forecast errors. Our results agree with those of Clements for output growth and inflation at long horizons, but differ at short horizons. One reason for the difference is that Clements uses point forecasts while expression (16) only holds for the mean: if absolute forecast errors are computed using predictions other than the mean, there is no *a priori* reason why there should be a correspondence with the subjective standard deviation, even under RE. In

[19] Clements adjusts for the unbalanced nature of the sample—that is, the fact that each forecaster's average is computed for a different time period—by constructing weighted averages where the weights reflect the average forecast error or subjective uncertainty during that period.

fact Figure A-20 shows that when we use the point predictions the correspondence between subjective uncertainty and forecast error vanishes at short horizons.[20]

The purely cross-sectional comparison undertaken by Clements (2014) misses the time dimension of our regression, where we investigate whether changes in subjective uncertainty over time actually map into changes in forecasting performance. This aspect is particularly important as it sheds light on whether forecasters correctly anticipate periods of macroeconomic uncertainty. For long horizons, they clearly do not. But as the horizon gets shorter, their assessment of uncertainty becomes more and more in line with the noisy RE model.

Last, one benefit of our approach is that we can measure inference uncertainty about $\sigma_{t|t-q,i}$. We can therefore assess to what extent such uncertainty may be driving the results in Figure 7. We do so by running a weighted OLS panel regression, where the weights are inversely proportional to inference uncertainty as measured by the interquantile range of the posterior distribution of $\sigma_{t|t-q,i}$. Figure 9 shows that the weighted OLS results are nearly identical to the results in Figure 7, assuaging concerns of an attenuation bias driven by inference uncertainty for long horizons. The appendix shows that the weighted results are very similar to the unweighted ones also whenever we use fixed effects and/or different samples.

III.C.3 A location test: The relative accuracy of mean and point predictions

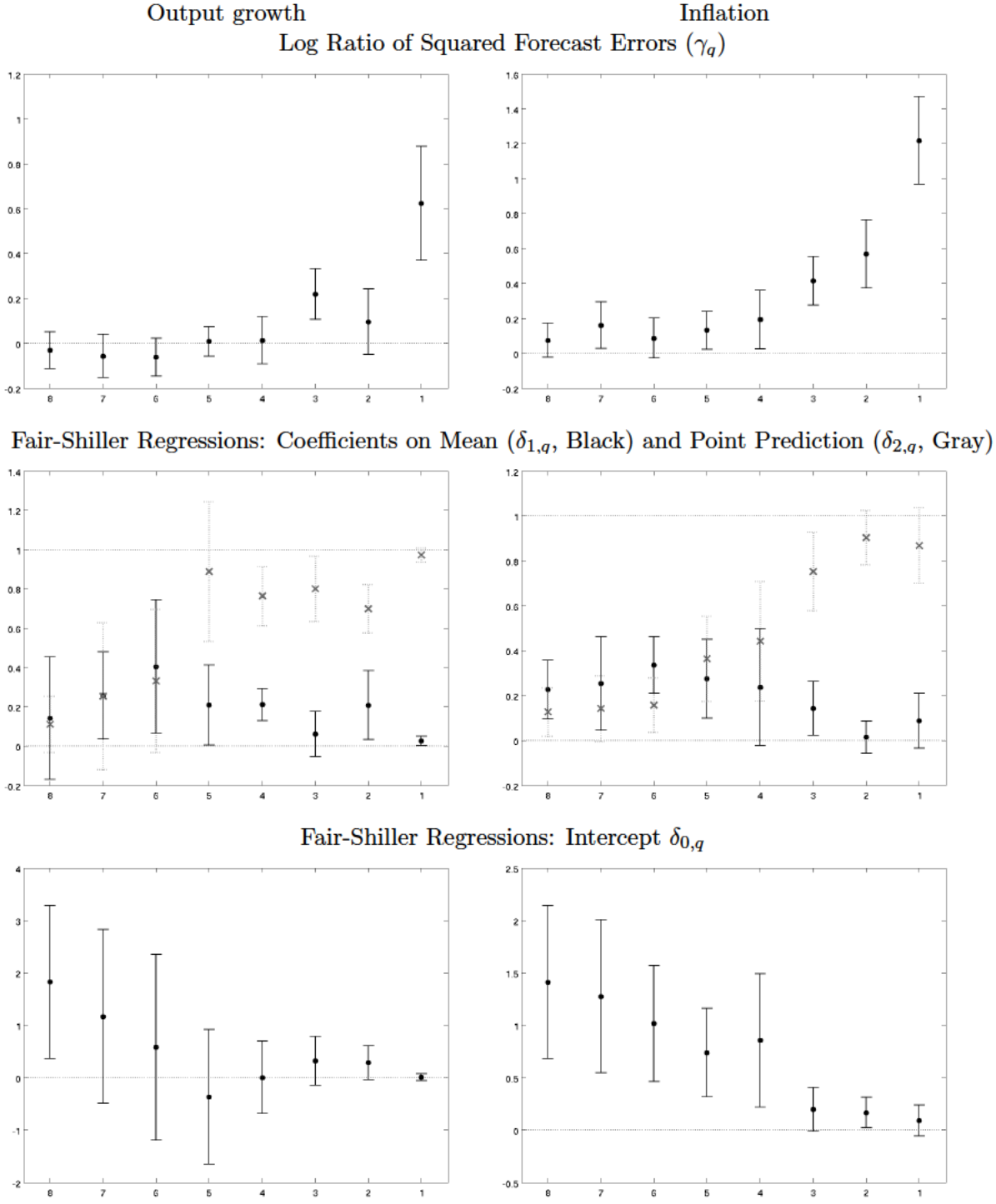
Finally, we turn to the *location* test, where we use the point forecast as an alternative to the mean projections in testing (17). The top panels of Figure 10 show OLS estimates of the coefficient γ_q in the panel regression

$$\ln \frac{(y_t - \mathbb{E}_{t-q,i}[y_t])^2}{(y_t - y^{pp}_{t,t-q,i})^2} = \gamma_q + \epsilon_{t,i,q}, \;\; t = 1, ..., T, \;\; i = 1, ..., n, \tag{20}$$

where $y^{pp}_{t,t-q,i}$ is the point forecast for y_t made by forecaster i in period $t-q$. Estimates of γ_q significantly greater than zero indicate that, on average, mean projections fare worse than point forecasts in terms of mean squared error. In fact, these estimates can be interpreted as the percentage improvement/worsening in forecast accuracy for point forecasts relative to mean projections.

[20] Figure A-21 shows that with time-fixed effects most coefficients are not significantly different from 0 for point forecasts. Figure A-22 shows the results obtained using the generalized beta approach, which have the same pattern as those shown in Figure 7 but are quantitatively different for short horizons.

Figure 10: A location test: Relative accuracy of mean vs. point projections



Note: Top panel: Black dots correspond to OLS estimates of γ_q from regression (20) for $q = 8, ..., 1$. Middle panel: Black dots and gray crosses correspond to OLS estimates of $\delta_{1,q}$ and $\delta_{2,q}$, respectively, from regression (21) for $q = 8, ..., 1$. Bottom panel: Black dots correspond to OLS estimates of the constant $\delta_{0,q}$ from regression (21) for $q = 8, ..., 1$. In all panels, whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors. Table A-6 in the appendix provides the regression statistics.

For horizons longer than one year, estimates of γ_q are not significantly different from zero for output growth and only slightly positive for inflation. This result may partly reflect the fact that, for these horizons, point and mean predictions are not very different (see Engelberg et al., 2009). As the horizon gets shorter, the estimates tend to become much larger and significantly positive for both output growth and inflation. The result that point forecasts perform better than mean forecasts in terms of mean squared error for short horizons is not new to the literature: Clements (2010) reports mean squared forecast errors for horizons shorter than one year and finds that these are lower for point than for mean projections. As in Clements (2010), we interpret these results explicitly as an indirect test of the rationality of density projections: under RE, it better be that the mean of the predictive distribution produces a lower mean squared error than any other point prediction regardless of the forecasters' loss function. The fact that for short horizons this is clearly not the case casts some doubt on explanations for the divergence between mean and point forecasts that rely on the forecasters' loss function (e.g., Patton and Timmermann, 2007; Elliott et al., 2008; Lahiri and Liu, 2009).

As a further test of the rationality of mean projections, we also run the Fair and Shiller (1990) regression

$$y_t = \delta_{0,q} + \delta_{1,q}\mathbb{E}_{t-q,i}[y_t] + \delta_{2,q}y^{pp}_{t,t-q,i} + \epsilon_{t,i,q},\ t = 1, ..., T,\ i = 1, ..., n. \tag{21}$$

The rationality of density projections would imply $\delta_{0,q} = 0$, $\delta_{1,q} = 1$, and $\delta_{2,q} = 0$. If point projections $y^{pp}_{t,t-q,i}$ coincide with mean forecasts, then the two regressors are collinear. The middle panels of Figure 10 report estimates of $\delta_{1,q}$ (black crosses) and $\delta_{2,q}$ (gray diamonds) for different horizons q, while the bottom panels report estimates for the constant $\delta_{0,q}$. Estimates of $\delta_{1,q}$ are significantly below 1 for all horizons, for both output and inflation. Estimates for the constant are significantly different from zero for long horizons for inflation, but only for $q = 8$ for output. As the horizon shortens, estimates for the constant become closer to 0 for both output and inflation, but estimates of $\delta_{2,q}$ rise toward 1 while estimates of $\delta_{1,q}$ fall to zero, indicating that point predictions are much closer to actual outcomes than mean forecasts.

III.C.4 Summing up: Are SPF density forecasts consistent with the noisy rational expectation hypothesis?

The body of evidence collected in this section suggests that the answer is no, in line with the results of the literature using only point forecasts (e.g., Bordalo et al., 2020; Angeletos et al.,

2021).[21] The evidence from second moments indicates that deviations from rationality are especially strong for longer horizon forecasts: 1) forecasters tend to be overconfident, and 2) there is little relationship between differences in subjective uncertainty across forecasters and over time and differences in forecasting performance. If the first finding comes with the caveat that the degree of overconfidence is very imprecisely estimated, especially for inflation, the second finding is statistically significant for both output and inflation. For short horizons, there is (almost) a one-to-one mapping between subjective and ex-post uncertainty, across forecasters and over time and for both variables, in accordance with the noisy RE hypothesis. Density forecasts are slightly underconfident for both output growth and inflation, however. But while the second moments of the density projections broadly line up with theory at short horizons, the first moments do not: mean projections deliver higher mean squared errors than point projections.

Our evidence that deviations from rationality vary with the horizon echoes the findings of some of the literature using point forecasts.[22] Angeletos et al. (2021), for instance, argue that forecasters overreact to shocks in the sense that they overestimate their persistence. This overreaction over longer horizons is arguably consistent with the results in Figure 7: a change in the perceived variance of shocks, over time and/or across forecasters, leads to a correct assessment of the change in subjective uncertainty in the short run, but a spurious change in subjective uncertainty as the horizon increases since forecasters tend to exaggerate the effect of this change. This longer-run overreaction may seem at odds with our finding that forecasters are overconfident for long horizons (Figure 5), however: if the forecasters' perceived response of the economy to future shocks is stronger than the actual response, one would expect them to overestimate uncertainty, which is the opposite of what we find. But uncertainty is due not only to the effect of shocks but also to model/parameter uncertainty. If forecasters use the incorrect model/parameter values but fail to recognize the impact of model/parameter uncertainty, they ignore the effect of the bias induced by misspecification on their forecast accuracy. The fact that this bias increases with the horizon, as shown in Figure 10, can therefore reconcile our findings with the theory and the evidence in Angeletos

[21] Valchev and Gemmi (2023) contend that deviations from rationality for point forecasts are due to strategic incentives rather than lack of rationality. Arguably the case for strategic incentives is harder to make for density forecasts, since these are anonymous, are not part of any public competition (e.g., Blue Chip), and are more difficult to rank.

[22] A small but interesting survey conducted by the Philadelphia Fed (Stark, 2013) indicates that SPF respondents use a combination of models and judgment to produce forecasts, but rely more on the former for short horizon projections and *vice versa* for longer horizon ones.

et al. (2021).

Patton and Timmermann (2010) show that dispersion in point forecasts increases with the horizon and argue that this result is consistent with a framework where forecasters have not only different information sets but also different priors and these priors matter more for longer horizons. Their evidence for point projections complements our finding that differences in density forecasts across forecasters and over time cannot be explained only by differences in the information set, especially for longer horizons.

III.C.5 Robustness

The consistency results of section II.E do not imply that our results are invariant to the choice of priors and functional forms given that our sample is finite and these results do not apply to some of the objects of interest. In the appendix we therefore reproduce all the figures in the paper under three different robustness checks: 1) halving the *a priori* noise, including the likelihood of reporting zeros (section E.A); 2) doubling the *a priori* noise, including the likelihood of reporting zeros (section E.B); and 3) using a mixture of three normals as our parametric choice for the $F(\cdot)$ function (section E.C), as opposed to a mixture of two as in our baseline results. We find that the main results in the paper—and in particular those pertaining to the degree on over- and underconfidence and the estimates for the variation test—are broadly robust, even though the inference for specific forecasters changes relative to our baseline assumptions (e.g., inference for the forecaster placing 80 percent mass on the open bin in Figure 2). Finally, section E.D shows the results under outlier Winsorization to make sure they are not driven by outliers. These results are virtually identical to those reported in the main text.

IV Conclusions

In this paper we presented a novel approach for conducting inference on data from probabilistic surveys and used it to investigate whether U.S. Survey of Professional Forecasters density projections for output growth and inflation are consistent with the noisy rational expectations hypothesis. We find that for horizons of close to two years there is no correspondence between subjective uncertainty and forecast accuracy for output growth density projections, both across forecasters and over time, and only a mild correspondence for inflation projections, in contrast to what rational expectations would predict. As the horizons

shorten, the relationship becomes one-to-one, in accordance with the theory. We also find that forecasters tend to be overconfident for long horizons, but underconfident for short horizons, although the RE benchmark is never rejected for inflation.

While the inference approach we propose arguably offers several benefits relative to current practice, it is important to point out some of its limitations. We provided consistency results that take advantage of the nonparametric nature of the approach, but these only apply to the model as a data-generating process for the data that we observe—the bin probabilities. Consistency results for the underlying continuous predictive densities are only available in the unrealistic case that the number of bins goes to infinity and the bin width goes to zero. When these conditions are not met, the results obtained with our approach may be sensitive to the choice of the base function and of priors, even when the number of forecasters goes to infinity. In addition, the approach proposed in this paper 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. We leave this extension to future research. Our tests of rationality for density forecasts also call for further work. Our findings speak to the average behavior of forecasters, in the spirit of much of the literature using point forecasts, but are silent on many potentially interesting sources of forecasters' heterogeneity.

References

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

Andrieu, Christophe and Johannes Thoms, "A tutorial on adaptive MCMC," *Statistics and Computing*, Dec 2008, *18* (4), 343–373.

Angeletos, George-Marios, Zhen Huo, and Karthik A Sastry, "Imperfect macroeconomic expectations: Evidence and theory," *NBER Macroeconomics Annual*, 2021, *35* (1), 1–86.

Barrientos, Andrés F., Alejandro Jara, and Fernando A. Quintana, "Bayesian density estimation for compositional data using random Bernstein polynomials," *Journal of Statistical Planning and Inference*, 2015, *166*, 116–125.

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.

_ , _ , and Marco Del Negro, "Inference on Probabilistic Surveys in Macroeconomics with an Application to the Evolution of Uncertainty in the Survey of Professional Forecasters during the COVID Pandemic," in "Handbook of Economic Expectations," Elsevier, 2023, pp. 443–476.

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

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 _ , "The Measurement and Charateristics of Professional Forecasts' Uncertainty," *Journal of Applied Econometrics*, 2014, *7* (30), 10291046.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer, "Overreaction in macroeconomic expectations," *American Economic Review*, 2020, *110* (9), 2748–82.

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.

Burda, Martin, Matthew Harding, and Jerry Hausman, "A Bayesian semiparametric competing risk model with unobserved heterogeneity," *Journal of Applied Econometrics*, 2015, *30* (3), 353–376.

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

Chib, Siddhartha and Barton H. Hamilton, "Semiparametric Bayes analysis of longitudinal data treatment models," *Journal of Econometrics*, 2002, *110* (1), 67–89.

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

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

_ **, Robert W. Rich, and Joseph S. Tracy**, "Chapter 3 - Surveys of professionals," in Rüdiger Bachmann, Giorgio Topa, and Wilbert van der Klaauw, eds., *Handbook of Economic Expectations*, Academic Press, 2023, pp. 71–106.

Coibion, Olivier and Yuriy Gorodnichenko, "What can survey forecasts tell us about information rigidities?," *Journal of Political Economy*, 2012, *120* (1), 116–159.

_ **and** _ , "Information rigidity and the expectations formation process: A simple framework and new facts," *American Economic Review*, 2015, *105* (8), 2644–78.

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.

Diebold, Francis X, Anthony S Tay, and Kenneth F Wallis, "Evaluating density forecasts of inflation: the Survey of Professional Forecasters," in R.F. Engle and H. White, eds., *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive WJ Granger*, Oxford University Press, 1999.

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

Driscoll, John C and Aart C Kraay, "Consistent covariance matrix estimation with spatially dependent panel data," *Review of Economics and Statistics*, 1998, *80* (4), 549–560.

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

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.

Fair, Ray C and Robert J Shiller, "Comparing information in forecasts from econometric models," *The American Economic Review*, 1990, pp. 375–389.

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 Multihorizon Uncertainty from Survey Density Forecasts," *Journal of Money, Credit and Banking*, 2020.

_ **, Elmar Mertens, and Todd E Clark**, "What Is the Predictive Value of SPF Point and Density Forecasts?," 2023.

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

Ghoshal, S., J. K. Gosh, and R. V. Ramamoorthi, "Consistent semiparametric Bayesian inference about a location parameter," *Journal of Statistical Planning and Inference*, 1999, *77* (2), 181–193.

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.

Glas, Alexander and Matthias Hartmann, "Uncertainty Measures from Partially Rounded Probabilistic Forecast Surveys," *Quantitative Economics*, 2022, *13*, 979 – 1022.

Griffin, Jim E., "Inference in infinite superpositions of non-Gaussian Ornstein–Uhlenbeck processes using Bayesian nonparametic methods," *Journal of Financial Econometrics*, 2011, *9* (3), 519–549.

_ **and Mark F.J. Steel**, "Semiparametric Bayesian inference for stochastic frontier models," *Journal of Econometrics*, 2004, *123* (1), 121–152.

_ , **Fernando Quintana, and Mark F.J. Steel**, "Flexible and Nonparametric Modeling," in John Geweke, Gary Koop, and Herman K. van Dijk, eds., *Handbook of Bayesian Econometrics*, Oxford University Press, 2011.

Gu, Jiaying and Roger Koenker, "Unobserved heterogeneity in income dynamics: An empirical Bayes perspective," *Journal of Business & Economic Statistics*, 2017, *35* (1), 1–16.

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

Jensen, Mark J. and John M. Maheu, "Bayesian semiparametric stochastic volatility modeling," *Journal of Econometrics*, 2010, *157* (2), 306–316.

Kalli, Maria, Jim E. Griffin, and Stephen G. Walker, "Slice sampling mixture models," *Statistics and Computing*, 2011, *21*, 93–105.

Kozeniauskas, Nicholas, Anna Orlik, and Laura Veldkamp, "What are uncertainty shocks?," *Journal of Monetary Economics*, 2018, *100*, 1–15.

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.

_ **and** _ , "On the use of density forecasts to identify asymmetry in forecasters' loss function," *Business avd Economic Statistics Section-JSM*, 2009, pp. 2396–2408.

Liu, Laura, "Density forecasts in panel data models: A semiparametric Bayesian perspective," *Journal of Business & Economic Statistics*, 2023, (41), 349–363.

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 Michael 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.

Norets, Andriy and Justinas Pelenis, "Bayesian modeling of joint and conditional distributions," *Journal of Econometrics*, 2012, *168* (332-346).

_ **and** _ , "Posterior consistency in conditional density estimation by covariate dependent mixtures.," *Econometric Theory*, 2014, *30* (3), 606–646.

Patton, Andrew J and Allan Timmermann, "Properties of optimal forecasts under asymmetric loss and nonlinearity," *Journal of Econometrics*, 2007, *140* (2), 884–918.

_ **and** _ , "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.

Pelenis, Justinas, "Bayesian regression with heteroscedastic error density and parametric mean function," *Journal of Econometrics*, 2014, *178*, 624–638.

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

Potter, Simon, "The advantages of probabilistic survey questions: remarks at the IT Forum and RCEA Bayesian Workshop, keynote address, Rimini, Italy, May 2016," Technical Report, Federal Reserve Bank of New York 2016.

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.

Rossi, Barbara and Tatevik Sekhposyan, "Macroeconomic Uncertainty Indices Based on Nowcast and Forecast Error Distributions," *American Economic Review*, May 2015, *105* (5), 650–55.

_ **and** _ , "Alternative tests for correct specification of conditional predictive densities," *Journal of Econometrics*, 2019, *208* (2), 638–657.

Sethuraman, Jayaram, "A constructive definition of Dirichlet priors," *Statistica Sinica*, 1994, *4*, 639–650.

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.

Valchev, Rosen and Luca Gemmi, "Biased Surveys," Technical Report, National Bureau of Economic Research 2023.

Walker, Stephen G., "Sampling the Dirichlet mixture model with slices," *Communications in Statistics – Simulation and Computation*, 2007, *36*, 45–54.

Wu, Yuefeng and Subhashis Ghosal, "Correction to: "Kullback Leibler property of kernel mixture priors in Bayesian density estimation"," *Electron. J. Stat.*, 2009, *3*, 316–317.

_ **and** _ , "Kullback Leibler property of kernel mixture priors in Bayesian density estimation," *Electron. J. Stat.*, 2009, *2*, 298–331.

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

Appendix

A Data description

We focus on the Survey of Professional Forecasters, managed since 1992 by the Federal Reserve Bank of Philadelphia, and previously by the American Statistical Association and the National Bureau of Economic Research. The panel of forecasters include university professors and private-sector macroeconomic researchers, and the composition of the panel changes gradually over time. The survey, which is performed quarterly, is mailed to panel members the day after the government release of quarterly data on the national income and product accounts. We restrict our attention to the two variables for which the SPF has probabilistic questions for a long enough time span, namely year-over-year GDP growth and GDP deflator inflation for the current and the following year over the sample 1982Q1-2022Q4.

B The Gibbs Sampler

The infinite mixture model is

$$h_G(\mathbf{z}) = \int h(\mathbf{z}|\boldsymbol{\theta})G(d\boldsymbol{\theta}) = \sum_{k=1}^{\infty} w_k h(\mathbf{z}|\boldsymbol{\theta}_k^*). \tag{A-1}$$

Our Gibbs sampler applied to the cross section of $\mathbf{z}_i$, $i = 1, \ldots, n$ uses the convenient approach proposed by Walker (2007) and Kalli et al. (2011). For each forecaster i, conditional on the sequence of weights w_k's ($w_{1:\infty}$) and the sequence of atoms $\boldsymbol{\theta}_k^*$'s ($\boldsymbol{\theta}_{1:\infty}$), expression (A-1) can be written as the marginal distribution of

$$h(\mathbf{z}_i, u_i|w_{1:\infty}, \boldsymbol{\theta}_{1:\infty}^*) = \sum_{k=1}^{\infty} \mathbb{I}(u_i < w_k)h(\mathbf{z}_i|\boldsymbol{\theta}_k^*) \tag{A-2}$$

with respect to u_i, where u_i is uniformly distributed over the interval $[0, 1]$, and independent across i, and $\mathbb{I}(\cdot)$ is an indicator function. This implies that the conditional distribution of $\mathbf{z}_i$ given u_i, the weights and the atoms, is

$$h(\mathbf{z}_i|u_i, w_{1:\infty}, \boldsymbol{\theta}_{1:\infty}^*) = \frac{1}{h(u_i|w_{1:\infty})} \sum_{k \in A(u_i|w_{1:\infty})} h(\mathbf{z}_i|\boldsymbol{\theta}_k^*), \tag{A-3}$$

where the set $A(u_i|w_{1:\infty})$ includes all the atoms with a weight w_k larger than u_i ($A(u_i|w_{1:\infty}) = \{k : u_i < w_k\}$), and the marginal $h(u_i|w_{1:\infty}) = \sum_{k=1}^{+\infty} \mathbb{I}(u_i < w_k)$ since each $h(\cdot|\boldsymbol{\theta}^*_k)$ integrates to one. Unlike expression (A-1), expression (A-3) is a *finite* mixture where each component has probability $\frac{1}{h(u_i|w_{1:\infty})}$, which is straightforward to draw from using standard methods. Specifically, we will use the auxiliary indicators d_i's, which are equal to k if we draw from the k^{th} mixture component (note that, given u_i, the k^{th} component will only be drawn if it belongs to the set $A(u_i|w_{1:\infty})$). The resulting complete-data likelihood function is

$$L(\mathbf{z}_{1:n}|u_{1:n}, d_{1:n}, v_{1:\infty}, \boldsymbol{\theta}_{1:\infty}) = \prod_{i=1}^{n} \mathbb{I}_{\{u_i < w_{d_i}\}} h(\mathbf{z}_i|\boldsymbol{\theta}^*_{d_i}) \tag{A-4}$$

with $d_i \in \{k : u_i < w_k\}$, where $v_{1:\infty}$ is the infinite dimensional sequence containing the stick-breaking components which map into the weights via expression (10).

Let $\mathcal{D}_k = \{i : d_i = k\}$ denote the set of indexes of the observations allocated to the k-th component of the mixture. Let $\mathcal{D} = \{k : \mathcal{D}_k \neq \emptyset\}$ denote the set of indexes of the non-empty mixture components (in the sense that at least one i is using the k^{th} component) and $\bar{d} = \max \mathcal{D}$ the overall number of stick-breaking components used. The Gibbs sampler works as follows:

1. $v_{1:\infty}, u_{1:n}|d_{1:n}, \boldsymbol{\theta}^*_{1:\infty}, \psi, \mathbf{z}_{1:n}$.
   Call $v_{1:\bar{d}}$ the stick-breaking elements associated with the mixture components that are being used (conditional on $d_{1:n}$). Following Kalli et al. (2011), drawing from the joint posterior of $v_{1:\bar{d}}$, $v_{\bar{d}+1:\infty}$, and $u_{1:n}$, conditional on all other parameters, is accomplished by sampling sequentially from: (a) the marginal distribution of $v_{1:\bar{d}}$, (b) the conditional distribution of $u_{1:n}$ given $v_{1:\bar{d}}$, and (c) from the conditional distribution of $v_{\bar{d}+1:\infty}$ given $u_{1:n}$ and $v_{1:\bar{d}}$.

   (a) $v_{1:\bar{d}}|d_{1:n}, \boldsymbol{\theta}^*_{1:\infty}, \psi, \mathbf{z}_{1:n}$.
   After integrating out the u_i's, the posterior of $v_{1:\infty}$ is proportional to

$$\begin{aligned} p(v_{1:\infty}|d_{1:n}, \boldsymbol{\theta}^*_{1:\infty}, \psi, \mathbf{z}_{1:n}) &\propto \left(\prod_{i=1}^{n} w_{d_i} h(\mathbf{z}_i|\boldsymbol{\theta}^*_{d_i})\right) \left(\prod_{l=1}^{\infty} (1-v_l)^{\psi-1}\right) \\ &\propto \left(\prod_{i=1}^{n} \left(v_{d_i} \prod_{l=1}^{d_i-1} (1-v_l)\right) h(\mathbf{z}_i|\boldsymbol{\theta}^*_{d_i})\right) \left(\prod_{l=1}^{\infty} (1-v_l)^{\psi-1}\right). \end{aligned}$$

Now note that since $v_{\bar{d}+1:\infty}$ do not enter the likelihood (A-4) – that is, the term within the first parenthesis – they can be easily integrated out resulting in

$$p(v_{1:\bar{d}}|d_{1:n}, \boldsymbol{\theta}^*_{1:\infty}, \psi, \mathbf{z}_{1:n}) \propto \left(\prod_{i=1}^{n}\left(v_{d_i}\prod_{l=1}^{d_i-1}(1-v_l)\right)h(\mathbf{z}_i|\boldsymbol{\theta}^*_{d_i})\right)\left(\prod_{l=1}^{\bar{d}}(1-v_l)^{\psi-1}\right).$$

Therefore samples for $v_{1:\bar{d}}$ are obtained by drawing each v_k independently from

$$\pi(v_k|u_{1:n}, d_{1:n}, \dots) \propto (1-v_k)^{\psi+b_k-1}v_k^{a_k} \tag{A-5}$$

where $a_k = \sum_{i=1}^{n}\mathbb{I}(d_i = k)$ and $b_k = \sum_{i=1}^{n}\mathbb{I}(d_i > k)$, that is, v_k is drawn from a $Beta(a_k+1, b_k+\psi)$.

(b) $u_{1:n}|v_{1:\bar{d}}, d_{1:n}, \boldsymbol{\theta}^*_{1:\infty}, \psi, \mathbf{z}_{1:n}$.
The likelihood (A-4), seen as a function of each u_i, $i = 1, \dots, n$, is simply a uniform distribution over $[0, w_{d_i}]$. Hence

$$\pi(u_i|\dots) \propto \frac{1}{w_{d_i}}\mathbb{I}(u_i < w_{d_i}). \tag{A-6}$$

(c) $v_{\bar{d}+1:\infty}|u_{1:n}, v_{1:\bar{d}}, d_{1:n}, \boldsymbol{\theta}^*_{1:\infty}, \psi, \mathbf{z}_{1:n}$.
Again, $v_{\bar{d}+1:\infty}$ do not enter the likelihood (A-4), so samples from those v_k with $k > \bar{d}$ are simply obtained by drawing from the prior $Beta(1, \psi)$:

$$\pi(v_k|u_{1:n}, d_{1:n}, \dots) \propto (1-v_k)^{\psi-1}. \tag{A-7}$$

Of course, even if it is straightforward to execute, we do not want to generate an infinite number of draws. Fortunately we do not need to, as explained in Walker (2007). Inspection of (A-4) reveals that those mixtures for which $w_k < u_i$ will never be used, at least given the draw for u_i. Let $\bar{n}_i$ the smallest integer such that $\sum_{k=1}^{\bar{n}_i} w_k \geq 1-u_i$. Since by construction $\sum_{k=1}^{\infty} w_k = 1$, it must be that $\sum_{\bar{n}_i+1}^{\infty} w_k < u_i$ and therefore, a fortiori, $w_k < u_i$ for $k > \bar{n}_i$. Now define $\bar{n} = \max\{\bar{n}_i, i = 1, \dots, n\}$. Conditional on $u_{1:n}$, at most we will use $\bar{n}$ mixture components in the estimation. Hence we only need to draw $v_{\bar{d}+1:\bar{n}}$.

2. $\boldsymbol{\theta}^*_{1:\infty}|v_{1:\infty}, u_{1:n}, d_{1:n}, \psi, \mathbf{z}_{1:n}$.
For the same argument given above, we actually do not have to draw an infinite number of atoms, but only as many as they may possibly be used (at least given the current draw of $u_{1:n}$) – that is, at most $\bar{n}$. Note also that given the way the u_i's are drawn (from a uniform distribution over $[0, w_{d_i}]$), if $k \in \mathcal{D}$ then $k \leq \bar{n}$.

(a) For $k \in \mathcal{D}$ draws of $\boldsymbol{\theta}_k^*$ are obtained from

$$\pi(\boldsymbol{\theta}_k^* | \dots) \propto \left(\prod_{i \in \mathcal{D}_k} h(\mathbf{z}_i | \boldsymbol{\theta}_k^*) \right) G_0(\boldsymbol{\theta}_k^*). \tag{A-8}$$

Since the joint distribution is not tractable, samples have been generated by Adaptive Metropolis Hastings (AMH) proposed in Andrieu and Thoms (2008). More specifically, at the j-th iteration of the AMH for a parameter $\boldsymbol{\theta}^*$ of dimension p the proposal distribution is

$$\boldsymbol{\theta}^{new} \sim \mathcal{N}(\boldsymbol{\theta}^{(j-1)^*}, \Upsilon^{(j)}) \tag{A-9}$$

with covariance matrix $\Upsilon^{(j)} = \exp\{\xi^{(j)}\} I_p$ where $\xi^{(j)}$ is adapted over the iterations as follows

$$\xi^{(j)} = \xi^{(j-1)} + \gamma^{(j)}(\hat{\alpha}^{(j-1)} - \bar{\alpha}), \tag{A-10}$$

where $\bar{\alpha} = 0.3$ represents the desired level of acceptance probability, and $\hat{\alpha}^{(j-1)}$ is the previous iteration estimate of the acceptance probability (i.e. the acceptance rate). The diminishing adaptation condition is satisfied by choosing $\gamma^{(j)} = j^{(-a)}$. In the application we set $a = 0.7$.

(b) For $k \notin \mathcal{D},\ k \leq \bar{n}$ draws of $\boldsymbol{\theta}_k^*$ are obtained via independent sampling from the base measure G_0.

We therefore obtained a sequence of draws $\boldsymbol{\theta}_{1:\bar{n}}^*$, which we will use in the next Gibbs sampler step.

3. $d_{1:n} | v_{1:\infty}, u_{1:n}, \boldsymbol{\theta}_{1:\infty}^*, \psi, \mathbf{z}_{1:n}$
Draws for each d_i, $i = 1, \dots, n$, are obtained by sampling from a multinomial with weights proportional to

$$\pi(d_i | \dots) \propto \mathbb{I}(u_i < w_{d_i}) h(\mathbf{z}_i | \boldsymbol{\theta}_{d_i}^*) \tag{A-11}$$

with $d_i \in \{1, \dots, \bar{n}_i\}$. Note that in this draw we consider all possible mixture components from 1 to $\bar{n}_i$, not only those used so far (that is, those in $\mathcal{D}$). They will be drawn proportionally to their ability to fit of the data, as measured by $h(\mathbf{z}_i | \boldsymbol{\theta}_k^*)$.

C Theoretical results

In this Section we discuss various theoretical properties of our model. Subsection C.A develops in details the study of posterior consistency briefly discussed in Section II.E of the main text. Among other results, it contains a full derivation of Theorem 1 given in the main text, restated below as Theorem A.1 for the readers' convenience. Subsection C.B considers the behaviour of the model as the number of bins goes to infinity. Subsection C.C deals with the posterior consistency of the consensus distribution. Finally, Subsection C.D contains all the proofs.

In the main text, the vector $\boldsymbol{\theta} = (\mu, \mu_\delta, \sigma_1, \sigma_2, \omega, \phi, \epsilon)$ represents the collection of the model parameters. The sub-vector $(\mu, \mu_\delta, \sigma_1, \sigma_2, \omega)$ is used to parametrize $F(\cdot|\boldsymbol{\theta})$, ϕ controls the Dirichlet noise in (6) and ϵ drives the zero-bin probability, $\varrho_j(\boldsymbol{\theta}) = \varrho(\nu_j(\boldsymbol{\theta}), \epsilon)$ for $j = 1, \ldots, J$. In what follows, we do not necessarily assume that $F(\cdot|\boldsymbol{\theta})$ is a mixture of two Gaussian distributions nor we assume that $\boldsymbol{\varrho}(\boldsymbol{\theta})$ has the specific form given in (7). In the subsequent sections of the appendix, we address a more general scenario where $\theta \in \Theta \subset \mathbb{R}^m$, and both $F(\cdot|\boldsymbol{\theta})$ and $\varrho_j(\boldsymbol{\theta})$ represent general continuous functions with respect to the parameter θ. We also allow $\phi(\boldsymbol{\theta})$ to be a continuous function of $\boldsymbol{\theta}$, writing $\phi(\boldsymbol{\theta})$. To revert to the case discussed in the main text, it suffices to incorporate $\phi(\boldsymbol{\theta})$ in $\boldsymbol{\theta}$ and set $\phi(\boldsymbol{\theta}) = \phi$.

Let us recall that we consider the non parametric prior given by the random density on $\mathbf{z}$ defined by

$$h_G(\mathbf{z}) := \int h(\mathbf{z}|\boldsymbol{\theta}) G(d\boldsymbol{\theta}) = \sum_{k=1}^{\infty} w_k h(\mathbf{z}|\boldsymbol{\theta}_k^*) \tag{A-12}$$

where $G \sim \mathcal{DP}(\psi_0, G_0)$.

C.A Posterior consistency

Before stating some results on posterior consistency, let us recall the setting outlined in Section II.E of the main text. Given a prior Π on $\mathcal{H}$, the set of all possible data generating densities, the posterior $\Pi(\cdot|\mathbf{z}_1, ..., \mathbf{z}_n)$ is said to be *weakly consistent* at h_0 if for every i.i.d. sequence $\mathbf{z}_1, ..., \mathbf{z}_n$ of random vectors with common density h_0 the posterior probability $\Pi(U|\mathbf{z}_1, ..., \mathbf{z}_n)$ converges a.s. to 1 as $n \to +\infty$ for every weak neighborhood U of h_0. In our model, the prior is $\Pi(U) = P\{h_G \in U\}$, where h_G is defined (A-12).

We now describe the class of densities $\mathcal{H}$. Recall that $\mathbf{z}_{\boldsymbol{\xi}}$ is the collection of strictly positive elements of $\mathbf{z}$ and that $\boldsymbol{\xi}$ is a vector indexing these positive z_j's. The Lebesgue

measure on the sub-set of Δ^J identified by $\boldsymbol{\xi}$ is denoted by $\mathcal{L}_{\boldsymbol{\xi}}$ and a σ-finite measure on Δ^J can be defined as $\lambda(d\mathbf{z}) = \sum_{\boldsymbol{\xi}} \mathcal{L}_{\boldsymbol{\xi}}(d\mathbf{z}_{\boldsymbol{\xi}})$. The set of all possible data generating densities $\mathcal{H}$ is the set of all the densities $g(\mathbf{z}) = g(\mathbf{z}_{\boldsymbol{\xi}}, \boldsymbol{\xi})$ absolutely continuous with respect to λ. Given two densities h_0 and g in $\mathcal{H}$ the Kullback-Leibler divergence between h_0 and g is then defined as

$$KL(h_0, g) = \int_{\Delta^J} h_0(\mathbf{z}) \log\Big(\frac{h_0(\mathbf{z})}{g(\mathbf{z})}\Big)\lambda(d\mathbf{z}). \tag{A-13}$$

Call $\mathcal{M}^*$ the set of finite (but arbitrarily large) mixtures of densities (5) that define the parametric component of our model, and $\mathcal{H}_0^*$ the set of densities that can be approximated in the Kullback-Leibler sense by densities in $\mathcal{M}^*$, i.e.

$$\mathcal{H}_0^* = \{h_0 \text{ density w.r.t. } \lambda\text{: } \forall\, \epsilon > 0\ \exists\, g \in \mathcal{M}^* \text{ s.t. } KL(h_0, g) \leq \varepsilon\ \}.$$

We now state our main results, which contains Theorem 1 reported in the main text.

Theorem A.1. Assume that Θ is open subset of $\mathbb{R}^m$ for some m and that

$$\boldsymbol{\theta} \mapsto (\varrho_1(\boldsymbol{\theta}), ..., \varrho_J(\boldsymbol{\theta}), \phi(\boldsymbol{\theta})\nu_1(\boldsymbol{\theta}), ..., \phi(\boldsymbol{\theta})\nu_J(\boldsymbol{\theta}))$$

is a continuous function such that $\phi(\boldsymbol{\theta})\nu_j(\boldsymbol{\theta}) > 0$ and $0 < \varrho_j(\boldsymbol{\theta}) < 1$ for every $j = 1, ..., J$. If G_0 has full support, then the posterior is weakly consistent at any density h_0 in $\mathcal{H}_0^*$ such that

$$\int_{\Delta^J} \left|\log\left(\prod_{j:z_j>0} z_j\right)\right| h_0(\mathbf{z})\lambda(d\mathbf{z}) < +\infty. \tag{A-14}$$

Remark A.1. In order to obtain Theorem 1 of the main text, it suffices to take $\varrho_j(\boldsymbol{\theta}) = \varrho(\nu_j(\boldsymbol{\theta}), \epsilon)$ for $j = 1, \ldots, J$, $\boldsymbol{\varrho}(\boldsymbol{\theta})$ given by (7), incorporate $\phi(\boldsymbol{\theta})$ in $\boldsymbol{\theta}$ and set $\phi(\boldsymbol{\theta}) = \phi$.

If $\varrho_j(\boldsymbol{\theta}) = 0$ for $j = 1, \ldots, J$, i.e. forecasters give non-zero probability to each bin, $\mathcal{H}$ is the set densities (absolutely continuous respect to the Lebesgue measure) on Δ^J, that is with probability one $z_j > 0$ for all j. In this case, Kullback-Leibler divergence between two distribution h_0, g on Δ^J is easily defined as

$$KL(h_0, g) := \int_{\mathcal{Z}} h_0(\mathbf{z}) \log\Big(\frac{h_0(\mathbf{z})}{g(\mathbf{z})}\Big)d\mathbf{z}.$$

As a variant of the main theorem, we get a simpler result for the case in which $\varrho_j(\boldsymbol{\theta}) = 0$ for all $j = 1, \ldots, J$. In this case $\mathcal{M}^*$ is replaced by the set $\mathcal{M}$ of finite mixtures of

$$h(\mathbf{z}|\boldsymbol{\theta}) = \frac{\prod_{j=1}^J \Gamma(\phi(\boldsymbol{\theta})\nu_j(\boldsymbol{\theta}))}{\Gamma\Big(\sum_{j=1}^J \phi(\boldsymbol{\theta})\nu_j(\boldsymbol{\theta})\Big)} \prod_{j=1}^{J-1} z_j^{\phi(\boldsymbol{\theta})\nu_j(\boldsymbol{\theta})-1}\Big(1 - \sum_{j=1}^{J-1} z_j\Big)^{\phi(\boldsymbol{\theta})\nu_J(\boldsymbol{\theta})-1}.$$

and $\mathcal{H}_0^*$ by the set $\mathcal{H}_0$ of densities on Δ^J that can be approximated in the Kullback-Leibler sense by densities in $\mathcal{M}$, i.e.

$$\mathcal{H}_0 = \{h_0 \text{ density on } \Delta^J\colon \forall\, \epsilon > 0\ \exists\, g \in \mathcal{M} \text{ s.t. } KL(h_0, g) \leq \varepsilon\ \}.$$

Theorem A.2. Let Θ be an open subset of $\mathbb{R}^m$ for some m and $\varrho_j(\boldsymbol{\theta}) = 0$ for all $j = 1, \ldots, J$. Assume that $\boldsymbol{\theta} \mapsto (\phi(\boldsymbol{\theta})\nu_1(\boldsymbol{\theta}), \ldots, \phi(\boldsymbol{\theta})\nu_J(\boldsymbol{\theta}))$ is a continuous function on $\mathbb{R}_+^J$ such that $\phi(\boldsymbol{\theta})\nu_j(\boldsymbol{\theta}) > 0$ for every $j = 1, \ldots, J$. If G_0 has full support, then the posterior is weakly consistent at any density h_0 in $\mathcal{H}_0$ such that

$$\int_{\Delta^J} \left| \log \left(\prod_{j=1}^{J-1} z_j \Big(1 - \sum_{j=1}^{J-1} z_j\Big) \right) \right| h_0(\mathbf{z}) d\mathbf{z} < +\infty. \tag{A-15}$$

Remark A.2. If $\varrho_j(\boldsymbol{\theta}) = 0$ for all $j = 1, \ldots, J$, $\phi(\boldsymbol{\theta}) = \phi$, and a mixture of normal distributions is assumed for the subjective distribution, that is

$$F(y|\boldsymbol{\theta}) = \sum_{i=1}^{M} \omega_i \Phi(y|\mu_i, \sigma_i^2) \tag{A-16}$$

then the parameter vector is $\boldsymbol{\theta} = (\mu_1, \ldots, \mu_M, \sigma_1^2, \ldots, \sigma_M^2, \omega_1, \ldots, \omega_M, \phi)$. If G_0 has full support, then the posterior is weakly consistent at any h_0 in $\mathcal{H}_0$ satisfying (A-15). Indeed, in this case $(\phi\nu_1(\boldsymbol{\theta}), \ldots, \phi\nu_J(\boldsymbol{\theta}))$ is a continuous function on $\mathbb{R}_+^J$ and $\phi\nu_j(\boldsymbol{\theta}) > 0$ for every $j = 1, \ldots, J$.

The next Proposition gives some conditions ensuring that any continuous density function belongs to $\mathcal{H}_0$.

Proposition A.1. Assume $\varrho_j(\boldsymbol{\theta}) = 0$ for all $j = 1, \ldots, J$ and that $\boldsymbol{\theta} \mapsto (\phi(\boldsymbol{\theta})\nu_1(\boldsymbol{\theta}), \ldots, \phi(\boldsymbol{\theta})\nu_J(\boldsymbol{\theta}))$ is a continuous function on $\mathbb{R}_+^J$ such that $\phi(\boldsymbol{\theta})\nu_j(\boldsymbol{\theta}) > 0$ for every $j = 1, \ldots, J$. If for every $\boldsymbol{a} = (a_1, \ldots, a_J) \in [1, +\infty)^J$ and $\delta > 0$, there is $\boldsymbol{\theta}_\delta$ in Θ such that $\|\boldsymbol{a} - \boldsymbol{a}_\delta\|_\infty \leq \delta$ with $\boldsymbol{a}_\delta = \phi(\boldsymbol{\theta}_\delta)(\nu_1(\boldsymbol{\theta}_\delta), \ldots, \nu_J(\boldsymbol{\theta}_\delta))$, then any continuous density function on Δ^J belongs to $\mathcal{H}_0$.

Remark A.3. Note that combining Theorem A.2 and Proposition A.1 one gets that, under the assumptions of Proposition A.1, if G_0 has full support, then the posterior is weakly consistent at any h_0 which is continuous on Δ^J and satisfies (A-15). An example in which all the conditions of Proposition A.1 are met is the fully nonparametric case

$$F(y|\boldsymbol{\theta}) = \sum_{j=1}^{J} \varphi_j \mathbb{I}_{A_j}(y) \tag{A-17}$$

where $A_j = [y_j, +\infty)$, $j = 1, \ldots, J-1$, $A_J = [y^+, +\infty]$ and $\nu_j(\boldsymbol{\theta}) = \varphi_j$, $j = 1, \ldots, J$. Conditions in Proposition A.1 are satisfied also in the Gaussian mixture case of (A-16) with $M = J - 1$.

C.B Model properties

In this section, we present some properties which illustrate the flexibility of our nonparametric random histogram model. The behaviour of the model as the number of bins goes to infinity shows that our framework is theoretically sound since it can be used to approximate any subjective distribution when (2) holds.

Let $\mathbf{z}_i$, $i = 1, \ldots, n$ be i.i.d. samples from $h_G(\mathbf{z})$ and assume the forecasters never report zero probabilities (that is $\varrho_j(\boldsymbol{\theta}) = 0 \ \forall j$), then in expectation z_{ij} coincides with $\nu_j(\boldsymbol{\theta})$: $\mathbb{E}[z_{ij}|\boldsymbol{\theta}] = \nu_j(\boldsymbol{\theta})$. Expression (A-1) then implies that the distribution of each z_{ij}will be centered at the infinite mixture of the bin probabilities ν_j's implied by each mixture component $F(\cdot|\boldsymbol{\theta}_k)$:

$$\mathbb{E}\left[z_{ij}|G\right] = \sum_{k=1}^{\infty} w_k \nu_j(\boldsymbol{\theta}_k) = \sum_{k=1}^{\infty} w_k (F(y_j|\boldsymbol{\theta}_k) - F(y_{j-1}|\boldsymbol{\theta}_k)). \tag{A-18}$$

We show that our random histogram (prior) model converges to an infinite dimensional (prior) model approximating any subjective distribution in the topology of weak convergence. This flexibility implies that the nonparametric prior alleviates possible misspecification issues.

Introduce a latent Dirichlet process $Z_{i,\infty}(\cdot)|\boldsymbol{\theta}_i \sim \mathcal{DP}(\phi(\boldsymbol{\theta}_i), F(\cdot|\boldsymbol{\theta}_i))$ with parameters $\phi(\boldsymbol{\theta}_i)$ and $F(\cdot|\boldsymbol{\theta}_i)$, given $\boldsymbol{\theta}_i$ from G. This process defines a random measure on the observation space $\mathcal{Y}$ of the variable of interest (inflation), that is the support set of the subjective distribution $F(\cdot|\boldsymbol{\theta})$, and admits the equivalent stick breaking representation

$$Z_{i\infty}(y) = \sum_{j=1}^{\infty} w_{ij} \mathbb{I}\{y_{ij} \leq y\} \tag{A-19}$$

where y_{ij} $j = 1, 2, \ldots$ are i.i.d. random variables with common distribution $F(\cdot|\boldsymbol{\theta}_i)$ and w_{ij} $j = 1, 2, \ldots$ are obtained by using a sequence of i.i.d. $\mathcal{B}e(1, \phi(\boldsymbol{\theta}_i))$ random variables.

Proposition A.2. If $\varrho_j(\boldsymbol{\theta}) = 0$ for $j = 1, \ldots, J$, the Bayesian model

$$\mathbf{z}_i|G \overset{ind}{\sim} h_G(\mathbf{z}), \quad i = 1, \ldots, n$$
$$G \sim \mathcal{DP}(\psi, G_0)$$

where $\mathbf{z}_i = (z_{i1}, \ldots, z_{iJ})$ admits the following stochastic representation:

$$(z_{i1}, \ldots, z_{iJ}) := (Z_{i,\infty}(y_1), Z_{i,\infty}(y_2) - Z_{i,\infty}(y_1), \ldots, 1 - Z_{i,\infty}(y_{J-1})), \quad i = 1, \ldots, n$$
$$Z_{i,\infty} \overset{ind}{\sim} \mathcal{DP}(\phi(\boldsymbol{\theta}_i), F(\cdot|\boldsymbol{\theta}_i)), \quad i = 1, \ldots, n$$
$$\boldsymbol{\theta}_i \overset{i.i.d.}{\sim} G \quad i = 1, \ldots, n$$
$$G \sim \mathcal{DP}(\psi, G_0).$$

The previous Proposition suggests the following interpretation: given the true subjective probability distribution $F(\cdot|\boldsymbol{\theta}_i)$ of the i-th forecaster and its level of noise $\phi(\boldsymbol{\theta}_i)$, the forecaster reports the weights $(z_{i1}, \ldots, z_{iJ})$ corresponding to the increments of a "noisy" version $Z_{i,\infty}$ of $F(\cdot|\boldsymbol{\theta}_i)$. This "noisy" version is the CDF obtained by a Dirichlet process with base measure $F(\cdot|\boldsymbol{\theta}_i)$ and concentration parameter $\phi(\boldsymbol{\theta}_i)$.

By (A-19), the latent Dirichlet process $Z_{i,\infty}$ is a random discrete CDF with infinite number of discontinuity points. To exemplify we depict $Z_{i,\infty}$ by the red stepwise line in Figure A-1. Despite of its discreteness, the process $Z_{i,\infty}$ ensures that our *prior model* gives positive probability to any weak neighbourhood of any distribution defined on the support set of $F(\cdot|\boldsymbol{\theta}_i)$. A combination of Proposition A.2 and Theorem 3.2.4 of Ghosh and Ramamoorthi (2003) gives the following result.

Corollary A.1. Assume that $\mathcal{Y} \subset \mathbb{R}$ is the support set of $F(\cdot|\boldsymbol{\theta})$ for any $\boldsymbol{\theta}$. Let $F(\cdot)$ be a distribution function with support subset of $\mathcal{Y}$, then $P(\{Z_{i,\infty} \in U_F\}) > 0$ for any weak neighbourhood U_F of $F(\cdot)$.

The random process $Z_{i\infty}$ can be seen as the limit of the histograms $\mathbf{z}_i$ when the number of bins goes to infinity. To show this formally, we associate the random histogram $\mathbf{z}_i$ to a random CDF Z_{iJ}. For any J we consider the partition $\mathcal{P}_J = \{y_0^J = -\infty < y_1^J < \ldots < y_J^J = +\infty\}$ and define the following one-to-one mapping between $\mathbf{z}_i$ and the CDF Z_{iJ}. Without loss of generality, we assign to the middle point of each interval the bin probability mass, and account for the two open bins (first and last) by introducing two auxiliary points y_-^J, y_+^J, such that $-\infty < y_-^J < y_1 < y_{J-1} < y_+^J < +\infty$. With this position we define the process $Z_{i,J}(y)$ (black line in Figure A-1):

$$Z_{iJ}(y) = \begin{cases} 0 & \text{if } y < y_-^J \\ z_{i1} & \text{if } y_-^J \leq y < (y_1^J + y_2^J)/2 \\ z_{i1} + \cdots + z_{ij} & \text{if } y \in [(y_{j-1}^J + y_j^J)/2, (y_j^J + y_{j+1}^J)/2) \text{ for } 2 < j \leq J-2 \\ z_{i1} + \cdots + z_{i\,J-1} & \text{if } y \in [(y_{J-2}^J + y_{J-1}^J)/2, y_+^J) \\ 1 & \text{if } y \geq y_+^J \end{cases}$$

The next theorem shows that Z_{iJ} converges to $Z_{i,\infty}$ with probability one in the topology of the weak convergence. Moreover, under continuity assumptions, the asymptotic mean of Z_{iJ}, conditionally on $\boldsymbol{\theta}_i$, coincides with the true subjective distribution. Note that, conditionally on $\boldsymbol{\theta}_i$, the mean of $Z_{i,\infty}$ is the true subjective distribution, i.e. $\mathbb{E}[Z_{i,\infty}(\cdot)|\boldsymbol{\theta}_i] = F(\cdot|\boldsymbol{\theta}_i)$.

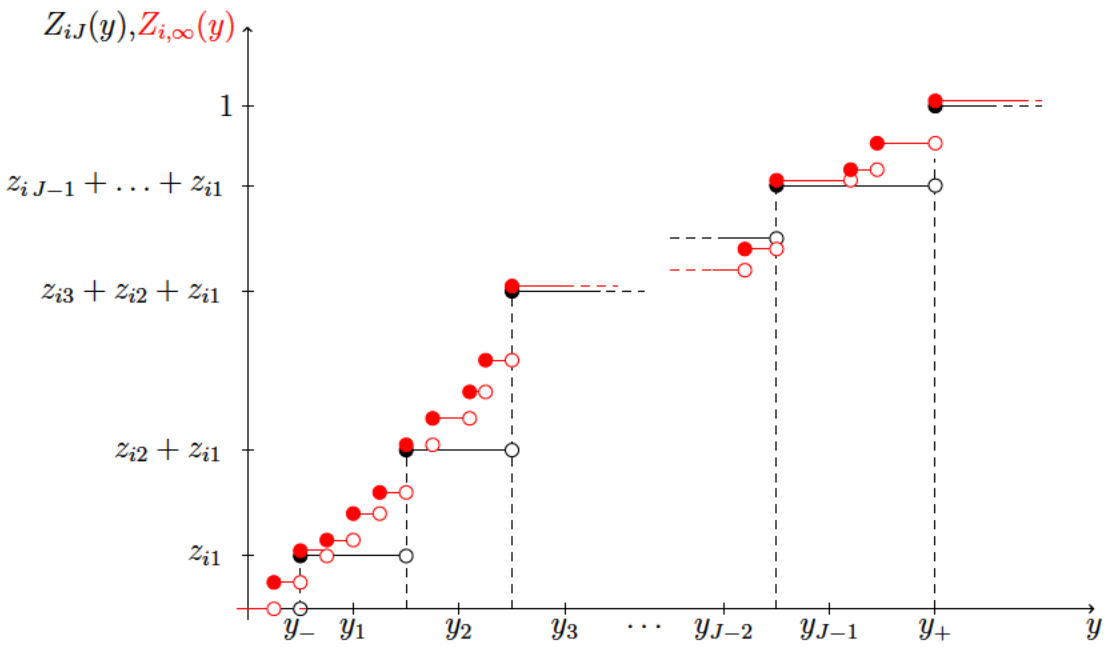


Figure A-1: Mapping between z_{ij}, $j = 1, \ldots, J$, $Z_{iJ}(y)$ and $Z_{i,\infty}(y)$.

Theorem A.3. Assume that $\varrho_j(\boldsymbol{\theta}) = 0$ for all j and the sequence of partitions $(\mathcal{P}_J)_J$ is such that $y_1 \to -\infty$, $y_{J-1} \to +\infty$ and $\max\{|y_{j+1} - y_j| : 1 \leq j \leq J-2\} \to 0$ for $J \to +\infty$. Then,

$$P\{\lim_{J\to+\infty} Z_{iJ}(y) = Z_{i,\infty}(y) \text{ for any } y \text{ point of continuity of } Z_{i,\infty}\} = 1.$$

If $F(\cdot|\boldsymbol{\theta}_i)$ is a continuous CDF, then

$$\lim_{J\to+\infty} \mathbb{E}[Z_{iJ}(y)|\boldsymbol{\theta}_i] = \mathbb{E}[Z_{i,\infty}(y)|\boldsymbol{\theta}_i] = F(y|\boldsymbol{\theta}_i) \qquad a.s.$$

C.C Consistency of the consensus distribution

The aggregate subjective distribution, also known as consensus distribution, is defined as

$$\bar{F}(y) = \frac{1}{n}\sum_{i=1}^{n} F_i(y)$$

where $F_i(y)$ is the forecast-specific subjective probability. In what follows, F_{n+1} denotes the posterior predictive distribution of y, defined as

$$F_{n+1}(y) := P\{y_{n+1} \leq y|\mathbf{z}_i, i = 1, \ldots, n\}.$$

The next proposition shows the connection between the two quantities in our model.

Proposition A.3. The distributions $\bar{F}$ and F_n are related by

$$F_{n+1}(y) = \frac{n}{n+\psi_0}\hat{F}_n(y) + \frac{\psi_0}{n+\psi_0}\int F(y|\theta)G_0(d\theta),$$

where $\hat{F}_n(y) = E(\bar{F}(y)|\mathbf{z}_1 ... \mathbf{z}_n)$

Using the previous relation one obtain a useful asymptotic properties of the consensus distribution.

Proposition A.4. Under the same assumptions of Theorem A.2,

$$\lim_{n\to+\infty}\left(F_{n+1}(y_i) - F_{n+1}(y_{i-1})\right) = \lim_{n\to+\infty}\left(\hat{F}_n(y_i) - \hat{F}_n(y_{i-1})\right) = \int z_i h_0(\mathbf{z})d\mathbf{z} \quad a.s.$$

for $i = 1, \ldots, J$. Hence, if there exists F^* such that $\int z_i h_0(\mathbf{z}) = F^*(y_i) - F^*(y_{i-1})$, then

$$\lim_{n\to+\infty} F_{n+1}(y_i) = \lim_{n\to+\infty} \hat{F}_n(y_i) = F^*(y_i) \quad a.s..$$

As in Subsection C.B, we consider set of nested partitions $\mathcal{P}_J = \{y_0^J = -\infty < y_1^J < \ldots < y_J^J = +\infty\}$ in such a way $\mathcal{P}_{J+1}$ is a refinement of $\mathcal{P}_J$. We assume that observations $\mathbf{z}_1^J, \ldots, \mathbf{z}_n^J$ are available with a "true" distribution $h_0 = h_0^J$ in $\mathcal{M}$, i.e. $h_0(\mathbf{z}) = \sum_{i=1}^{M} w_{i,0} h(\mathbf{z}|\boldsymbol{\theta}_{i,0})$ for suitable integer M, positive weights $(w_{1,0}, \ldots, w_{M,0})$ and parameters $\boldsymbol{\theta}_{1,0}, \ldots, \boldsymbol{\theta}_{M,0}$ in Θ. Note that with these hypotheses $\mathbf{z}_1^J, \ldots, \mathbf{z}_n^J$ are consistent in J, that is if $J' > J$ then $z_i^J = \sum_{j\in I(i)} z_j^{J'}$ if the i-th bin in $\mathcal{P}_J$ correspond the the union of the bins $j \in I(i)$ in $\mathcal{P}_{J'}$. This allows to consider limit jointly in the number of observations $(n \to +\infty)$ and in the number of bins $(J \to +\infty)$. Note also that for every J and every bin $(y_{i-1}, y_i]$ in $\mathcal{P}_J$

$$\int z_i h_0^J(\mathbf{z}) = F^*(y_i) - F^*(y_{i-1}),$$

for

$$F^*(y) := \sum_{i=1}^{M} w_{i,0} F(y|\boldsymbol{\theta}_{i,0}).$$

Proposition A.5. In the setting described above, under the same assumptions of Theorem A.3 on $\mathcal{P}_J$, then

$$\lim_{J\to+\infty, n\to+\infty} F_{n+1}(y) = \lim_{J\to+\infty, n\to+\infty} \hat{F}_n(y) = F^*(y) \quad a.s.$$

for every y point of continuity of F^*.

C.D Proofs

C.D.1 Proofs of Theorem A.1 and Theorem A.2

The proof of Theorem A.1 and Theorem A.2 are based on an application of Theorem 1 and Lemma 3 in Wu and Ghosal (2009b,a). In order to prove our theorems we need a slight generalization of these results. For the shake of clarity we state and prove this generalization.

In what follows, we denote with $supp(\mu)$ the weak support of a probability measure μ. We assume that $\mathcal{X}_0$ is a subset the finte set $\mathcal{X} = \{\boldsymbol{\xi} \in \{0,1\}^J : |\boldsymbol{\xi}| < J\}$. Let $\mathcal{Z}$ be the sample space, i.e. the set of all the pairs $(\boldsymbol{\xi}, \mathbf{z}_{\boldsymbol{\xi}})$ where $\boldsymbol{\xi} = (\xi_1, \dots, \xi_J)$, $\xi_i = \mathbb{I}\{z_i = 0\}$ and $\mathbf{z}_{\boldsymbol{\xi}}$ are the non-null elements of $\mathbf{z}$. In what follows we assume that $\mathbf{z}_{\boldsymbol{\xi}}$ takes values in an open subset $\mathcal{Z}_{\boldsymbol{\xi}}$ of $\mathbb{R}^{J-|\boldsymbol{\xi}|-1}$. In our application $\mathcal{Z}_{\boldsymbol{\xi}} = \Delta^{J-|\boldsymbol{\xi}|}$. On the sample space $\mathcal{Z}$, one defines the σ-finte measure $\lambda(d\mathbf{z}) = c(d\boldsymbol{\xi}) \otimes \mathcal{L}_{\boldsymbol{\xi}}(d\mathbf{z}_{\boldsymbol{\xi}})$ where c is the counting measure on $\mathcal{X}$ and, given $\boldsymbol{\xi}$, $\mathcal{L}_{\boldsymbol{\xi}}$ is the Lebesgue measure on $\mathcal{Z}_{\boldsymbol{\xi}} \subset \mathbb{R}^{J-|\boldsymbol{\xi}|-1}$. Let $\mathcal{H}$ be the set of all the densities with respect to λ and note that the densities g factorize as $g(\mathbf{z}) = g(\boldsymbol{\xi})g(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi})$. We also assume that the kernel $h(\mathbf{z}|\boldsymbol{\theta})$ factorizes in the same way, i.e.

$$h(\mathbf{z}|\boldsymbol{\theta}) = h(\boldsymbol{\xi}|\boldsymbol{\theta})h(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi}, \boldsymbol{\theta}).$$

Given a probability measure G on Θ, recall that

$$h_G(\mathbf{z}) = \int_\Theta h(\mathbf{z}|\boldsymbol{\theta})G(d\boldsymbol{\theta}). \tag{A-20}$$

Finally, assume that Π is the prior on $\mathcal{H}$ induced by the map (A-20) when G has prior $\hat{\Pi}$.

In our application, $h(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi}, \boldsymbol{\theta})$ is the pdf of a Dirichlet distribution, i.e. $\mathcal{D}ir(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\nu}(\boldsymbol{\theta}), \boldsymbol{\theta})$, $h_G(\mathbf{z})$ is given by (5) and $\hat{\Pi}$ is the Dirichlet process prior $\mathcal{DP}(\psi, G_0)$.

Given two densities h_0 and g in $\mathcal{H}$ the Kullback-Leibler divergence between h_0 and g is defined as

$$KL(h_0, g) = \int_{\mathcal{Z}} h_0(\mathbf{z}) \log\Big(\frac{h_0(\mathbf{z})}{g(\mathbf{z})}\Big)\lambda(d\mathbf{z}).$$

Hence, writing $h_0(\mathbf{z}) = h_0(\boldsymbol{\xi})h_0(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi})$ by Fubini Theorem one can re-arrange the previous expression as

$$\begin{aligned}
&\sum_{\boldsymbol{\xi}\in\mathcal{X}} h_0(\boldsymbol{\xi}) \int_{\mathcal{Z}_{\boldsymbol{\xi}}} h_0(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi}) \log\Big(\frac{h_0(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi})h_0(\boldsymbol{\xi})}{g(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi})g(\boldsymbol{\xi})}\Big)d\mathbf{z}_{\boldsymbol{\xi}} \\
&\qquad = \sum_{\boldsymbol{\xi}\in\mathcal{X}} h_0(\boldsymbol{\xi})\Big(\log\Big(\frac{h_0(\boldsymbol{\xi})}{g(\boldsymbol{\xi})}\Big) + \int_{\mathcal{Z}_{\boldsymbol{\xi}}} h_0(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi}) \log\Big(\frac{h_0(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi})}{g(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi})}\Big)d\mathbf{z}_{\boldsymbol{\xi}}\Big),
\end{aligned}$$

where for simplicity we write $\mathcal{L}_{\boldsymbol{\xi}}(d\mathbf{z}_{\boldsymbol{\xi}})$ simply as $d\mathbf{z}_{\boldsymbol{\xi}}$

Theorem A.4. Let Θ be a Polish space and h_0 a density in $\mathcal{H}$. If for any $\varepsilon > 0$ there is a probability measure $G_\varepsilon \in supp(\hat{\Pi})$ and a closed set D_ε in Θ such that

(H1) $KL(h_0, h_{G_\varepsilon}) = \sum_{\boldsymbol{\xi}\in\mathcal{X}} h_0(\boldsymbol{\xi}) \int_{\mathcal{Z}_{\boldsymbol{\xi}}} \log\Big(\frac{h_0(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi})h_0(\boldsymbol{\xi})}{h_{G_\varepsilon}(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi})h_{G_\varepsilon}(\boldsymbol{\xi})}\Big) h_0(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi},)d\mathbf{z}_{\boldsymbol{\xi}} < \varepsilon$;

(H2) D_ε contains $supp(G_\varepsilon)$ in its interior and for every $\boldsymbol{\xi}$

$$\int_{\mathcal{Z}_{\boldsymbol{\xi}}} \log\Big(\frac{h_{G_\varepsilon}(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi})h_{G_\varepsilon}(\boldsymbol{\xi})}{\inf_{\boldsymbol{\theta}\in D_\varepsilon} h(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi},\boldsymbol{\theta})h(\boldsymbol{\xi}|\boldsymbol{\theta})}\Big) h_0(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi})d\mathbf{z}_{\boldsymbol{\xi}} < +\infty;$$

(H3) $\inf_{\mathbf{z}_{\boldsymbol{\xi}}\in C_{\boldsymbol{\xi}}} \inf_{\boldsymbol{\theta}\in D_\varepsilon} h(\boldsymbol{\xi}|\boldsymbol{\theta})h(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi},\boldsymbol{\theta}) > 0$ for every $\boldsymbol{\xi}$ and every compact set $C_{\boldsymbol{\xi}}$ in $\mathcal{Z}_{\boldsymbol{\xi}}$;

(H4) $\{\boldsymbol{\theta} \mapsto h(\boldsymbol{\xi}|\boldsymbol{\theta})h(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi},\boldsymbol{\theta}) : \mathbf{z}_{\boldsymbol{\xi}} \in C_{\boldsymbol{\xi}}\}$ is uniformly equicontinuous on D_ε, for every $\boldsymbol{\xi}$ and every compact set $C_{\boldsymbol{\xi}}$ in $\mathcal{Z}_{\boldsymbol{\xi}}$;

then $\Pi\{KL(h_0, h_G) \geq \varepsilon\} > 0$ for every $\varepsilon > 0$ and hence Π is weakly consistent at h_0.

Assumption (H1) corresponds to (A1) in Theorem 1 of Wu and Ghosal (2009b). Assumptions (H2)-(H3) correspond to assumptions (A7)-(A8) of Lemma 3 of Wu and Ghosal (2009b), while (H4) is slightly different from the original assumption (A9), see Wu and Ghosal (2009a). The theorem reduces to Theorem 1 and Lemma 3 of Wu and Ghosal (2009b) when $\mathcal{X}_0$ is the single point $\boldsymbol{\xi} = (0, \ldots, 0)$.

Proof of Theorem A.4. One has

$$\begin{aligned} KL(h_0, h_G) &= KL(h_0, h_{G_\varepsilon}) + \sum_{\boldsymbol{\xi}\in\mathcal{X}} h_0(\boldsymbol{\xi}) \int_{\Delta^{J-|\boldsymbol{\xi}|}} \log\Big(\frac{h_{G_\varepsilon}(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi})h_{G_\varepsilon}(\boldsymbol{\xi})}{h_G(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi})h_G(\boldsymbol{\xi})}\Big) h_0(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi})d\mathbf{z}_{\boldsymbol{\xi}} \\ &\leq \varepsilon + \sum_{\boldsymbol{\xi}\in\mathcal{X}} h_0(\boldsymbol{\xi}) \int_{\Delta^{J-|\boldsymbol{\xi}|}} \log\Big(\frac{h_{G_\varepsilon}(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi})h_{G_\varepsilon}(\boldsymbol{\xi})}{h_G(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi})h_G(\boldsymbol{\xi})}\Big) h_0(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi})d\mathbf{z}_{\boldsymbol{\xi}} =: \varepsilon + A_\varepsilon(G). \end{aligned}$$

If we show that there is an open neighbourhood V of G_ε such that for every G in V one has $A_\varepsilon(G) \leq \varepsilon$, then $\Pi\{KL(h_0, h_G) \geq 2\varepsilon\} > 0$ for every $\varepsilon > 0$. To prove the claim, for every $\boldsymbol{\xi}$ by (H2) we find a compact set $C_{\boldsymbol{\xi}}$ such that

$$\int_{C_{\boldsymbol{\xi}}^c} \log\Big(\frac{h_{G_\varepsilon}(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi})h_{G_\varepsilon}(\boldsymbol{\xi})}{\inf_{\boldsymbol{\theta}\in D_\varepsilon} h(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi},\boldsymbol{\theta})h(\boldsymbol{\xi}|\boldsymbol{\theta})}\Big) h_0(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi})d\mathbf{z}_{\boldsymbol{\xi}} \leq \frac{\varepsilon}{4}$$

and

$$\int_{C_{\boldsymbol{\xi}}^c} h_0(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi})d\mathbf{z}_{\boldsymbol{\xi}} \leq \frac{\varepsilon}{4\log(2)}.$$

Let $V_0 := \{G : G(D_\varepsilon) > 1/2\}$. Since $G_\varepsilon(D_\varepsilon) = 1$, by Portmanteau Theorem V_0 is an open neighbourhood of G_ε. Now

$$h_G(\boldsymbol{\xi}, \mathbf{z}_{\boldsymbol{\xi}}) = \int_{D_\varepsilon} h(\boldsymbol{\xi}, \mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\theta}) G(d\boldsymbol{\theta}) \geq \inf_{\boldsymbol{\theta} \in D_\varepsilon} h(\boldsymbol{\xi}|\boldsymbol{\theta}) h(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi}, \boldsymbol{\theta}) G(D_\varepsilon),$$

hence, for every G in V_0,

$$\begin{aligned} &\int_{C_{\boldsymbol{\xi}}^c} \log\Big(\frac{h_{G_\varepsilon}(\boldsymbol{\xi}, \mathbf{z}_{\boldsymbol{\xi}})}{h_G(\boldsymbol{\xi}, \mathbf{z}_{\boldsymbol{\xi}})}\Big) h_0(\boldsymbol{\xi}, \mathbf{z}_{\boldsymbol{\xi}}) d\mathbf{z}_{\boldsymbol{\xi}} \\ &\quad \leq \int_{C_{\boldsymbol{\xi}}^c} \log\Big(\frac{h_{G_\varepsilon}(\boldsymbol{\xi}, \mathbf{z}_{\boldsymbol{\xi}})}{\inf_{\boldsymbol{\theta} \in D_\varepsilon} h(\boldsymbol{\xi}|\boldsymbol{\theta}) h(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi}, \boldsymbol{\theta})}\Big) h_0(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi}) d\mathbf{z}_{\boldsymbol{\xi}} + \log(2) \int_{C_{\boldsymbol{\xi}}^c} h_0(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi}) d\mathbf{z}_{\boldsymbol{\xi}} \leq \frac{\varepsilon}{2}. \end{aligned} \tag{A-21}$$

By condition (H4), for every $\boldsymbol{\xi}$ there are $\mathbf{z}_{\boldsymbol{\xi}}^{(i)} \in C_{\boldsymbol{\xi}}$ $i = 1, \ldots, m$, such that for every $\mathbf{z}_{\boldsymbol{\xi}} \in C_{\boldsymbol{\xi}}$ there is i for which

$$\sup_{\boldsymbol{\theta} \in D_\varepsilon} |h(\boldsymbol{\xi}|\boldsymbol{\theta}) h(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi}, \boldsymbol{\theta}) - h(\boldsymbol{\xi}, \mathbf{z}_{\boldsymbol{\xi}}^{(i)}|\boldsymbol{\theta})| \leq \frac{c\varepsilon}{12}$$

where $c := \inf_{\mathbf{z}_{\boldsymbol{\xi}} \in C_{\boldsymbol{\xi}}} \inf_{\boldsymbol{\theta} \in D_\varepsilon} h(\boldsymbol{\xi}|\boldsymbol{\theta}) h(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi}, \boldsymbol{\theta}) > 0$ by (H3). Since $G_\varepsilon(\partial D_\varepsilon) = 0$, the set

$$V_{\boldsymbol{\xi}} := \{G : \Big| \int_{D_\varepsilon} h(\boldsymbol{\xi}, \mathbf{z}_{\boldsymbol{\xi}}^{(i)}|\boldsymbol{\theta}) G_\varepsilon(d\boldsymbol{\theta}) - \int_{D_\varepsilon} h(\boldsymbol{\xi}, \mathbf{z}_{\boldsymbol{\xi}}^{(i)}|\boldsymbol{\theta}) G(d\boldsymbol{\theta}) \Big| < \frac{c\varepsilon}{12}; \quad i = 1, \ldots, m\}$$

is a weak neighbourhood of G_ε. Hence, for G in $V_{\boldsymbol{\xi}}$

$$\Big| \int_{D_\varepsilon} h(\boldsymbol{\xi}, \mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\theta}) G_\varepsilon(d\boldsymbol{\theta}) - \int_{D_\varepsilon} h(\boldsymbol{\xi}, \mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\theta}) G(d\boldsymbol{\theta}) \Big| \leq \frac{c\varepsilon}{4} \tag{A-22}$$

Since $supp(G_\varepsilon) \subset D_\varepsilon$,

$$\int_{C_{\boldsymbol{\xi}}} \log\Big(\frac{h_{G_\varepsilon}(\boldsymbol{\xi}, \mathbf{z}_{\boldsymbol{\xi}})}{h_G(\boldsymbol{\xi}, \mathbf{z}_{\boldsymbol{\xi}})}\Big) h_0(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi}) d\mathbf{z}_{\boldsymbol{\xi}} \leq \int_{C_{\boldsymbol{\xi}}} \log\Big(\frac{\int_{D_\varepsilon} h(\boldsymbol{\xi}, \mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\theta}) G_\varepsilon(d\boldsymbol{\theta})}{\int_{D_\varepsilon} h(\boldsymbol{\xi}, \mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\theta}) G(d\boldsymbol{\theta})}\Big) h_0(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi}) d\mathbf{z}_{\boldsymbol{\xi}}.$$

Hence, using $\log(x+1) \leq x$ and (A-22), for G in $V_0 \cap V_{\boldsymbol{\xi}}$ one obtains

$$\int_{C_{\boldsymbol{\xi}}} \log\Big(\frac{h_{G_\varepsilon}(\boldsymbol{\xi}, \mathbf{z}_{\boldsymbol{\xi}})}{h_G(\boldsymbol{\xi}, \mathbf{z}_{\boldsymbol{\xi}})}\Big) h_0(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi}) d\mathbf{z}_{\boldsymbol{\xi}} \leq \frac{\varepsilon}{2}. \tag{A-23}$$

At this stage, combining (A-21) and (A-23), one obtains that $A_\varepsilon(G) \leq \epsilon$ for every G in $V = V_0 \cap (\cap_{\boldsymbol{\xi}} V_{\boldsymbol{\xi}})$. □

We can now prove both Theorem A.1 and Theorem A.2.

Proof of Theorems A.1 and A.2. We start with the proof of Theorem A.2 because it is simpler from a notational point of view. We apply Theorem A.4 for $\mathcal{X}_0 = \{(0,\ldots,0)\}$. Let

$$\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}) := (\tilde{\nu}_1(\boldsymbol{\theta}),\ldots,\tilde{\nu}_J(\boldsymbol{\theta})) = (\phi(\boldsymbol{\theta})\nu_1(\boldsymbol{\theta}),\ldots,\phi(\boldsymbol{\theta})\nu_J(\boldsymbol{\theta})) \tag{A-24}$$

and

$$Z_{\boldsymbol{\theta}} = \frac{\prod_{j=1}^{J}\Gamma(\tilde{\nu}_j(\boldsymbol{\theta}))}{\Gamma\Big(\sum_{j=1}^{J}\tilde{\nu}_j(\boldsymbol{\theta})\Big)}.$$

Verification of (H1) of Theorem A.4. By hypothesis, for every $\varepsilon > 0$ there is $g_\varepsilon(\mathbf{z}) = \sum_{i=1}^{M_\varepsilon} w_{i,\varepsilon} h(\mathbf{z}|\boldsymbol{\theta}_{i,\varepsilon})$ in $\mathcal{M}$ such that $KL(h_0, g_\varepsilon) \leq \varepsilon$. To see that (H1) is satisfied, write $g_\varepsilon(\mathbf{z}) = \int h(\mathbf{z}|\boldsymbol{\theta}) G_\varepsilon(d\boldsymbol{\theta}) = h_{G_\varepsilon}(\mathbf{z})$ for $G_\varepsilon(d\boldsymbol{\theta}) = \sum_{i=1}^{M_\varepsilon} w_{i,\varepsilon}\delta_{\boldsymbol{\theta}_{i,\varepsilon}}(d\boldsymbol{\theta})$. Now $supp(G_\varepsilon) = \cup_{i=1}^{M_\varepsilon}\{\boldsymbol{\theta}_{i,\varepsilon}\}$. To conclude recall that if $\hat{\Pi}$ is $\mathcal{DP}(\psi, G_0)$ and $supp(G_\varepsilon) \subset supp(G_0)$, then $G_\varepsilon \in supp(\hat{\Pi})$; see, for instance, Theorem 3.2.4 of Ghosh and Ramamoorthi (2003).

Verification of (H2) of Theorem A.4. Given G_ε as above, one can find a compact set D_ε in Θ such that D_ε contains $\cup_{i=1}^{M_\varepsilon}\{\boldsymbol{\theta}_{i,\varepsilon}\} = supp(G_\varepsilon)$ in its interior.

Now

$$\begin{aligned} I_\varepsilon(\mathbf{z}) &:= \inf_{\boldsymbol{\theta}\in D_\varepsilon} h(\mathbf{z}|\boldsymbol{\theta}) \\ &= \inf_{\boldsymbol{\theta}\in D_\varepsilon} \frac{1}{Z_{\boldsymbol{\theta}}} \prod_{j=1}^{J-1} z_j^{\tilde{\nu}_j(\boldsymbol{\theta})-1}\Big(1-\sum_{j=1}^{J-1} z_j\Big)^{\tilde{\nu}_j(\boldsymbol{\theta})-1} \\ &\geq C_{1,\varepsilon} \prod_{j=1}^{J-1} z_j^{\mu_{j,\varepsilon}-1}\Big(1-\sum_{j=1}^{J-1} z_j\Big)^{\mu_{J,\varepsilon}-1} =: I^*_\varepsilon(\mathbf{z}), \end{aligned}$$

where $C_{1,\varepsilon} = \inf_{\boldsymbol{\theta}\in D_\varepsilon} Z_{\boldsymbol{\theta}}^{-1}$, $\mu_{j,\varepsilon} := \sup\{\tilde{\nu}_j(\boldsymbol{\theta}) : \boldsymbol{\theta} \in D_\varepsilon\}$. Now one has that $C_{1,\varepsilon} > 0$ and $\mu_{j,\varepsilon} > 0$, since D_ε is compact and the $\nu_j(\boldsymbol{\theta})$s are continuous and strictly positive.

On the one hand $h_{G_\varepsilon}(\mathbf{z}) \geq I_\varepsilon(\mathbf{z})$ and hence $\log(h_{G_\varepsilon}(\mathbf{z})/I_\varepsilon(\mathbf{z})) \geq 0$, on the other hand

$$\begin{aligned} \int \log\Big(\frac{h_{G_\varepsilon}(\mathbf{z})}{I_\varepsilon(\mathbf{z})}\Big) h_0(\mathbf{z})d\mathbf{z} &\leq \int \log\Big(\frac{h_{G_\varepsilon}(\mathbf{z})}{I^*_\varepsilon(\mathbf{z})}\Big) h_0(\mathbf{z})d\mathbf{z} \\ &\leq \int \left|\log\left(\frac{g_\varepsilon(\mathbf{z})}{\prod_{j=1}^{J-1} z_j^{\mu_{j,\varepsilon}-1}\Big(1-\sum_{j=1}^{J-1} z_j\Big)^{\mu_{J,\varepsilon}-1}}\right)\right| h_0(\mathbf{z})d\mathbf{z} + |\log(C_{1,\varepsilon})|. \end{aligned}$$

Since

$$C_{2,\varepsilon}\prod_{j=1}^{J-1} z_j^{A_{j,\varepsilon}-1}\Big(1-\sum_{j=1}^{J-1} z_j\Big)^{A_{J,\varepsilon}-1} \leq g_\varepsilon(\mathbf{z}) \leq C_{3,\varepsilon}\prod_{j=1}^{J-1} z_j^{B_{j,\varepsilon}-1}\Big(1-\sum_{j=1}^{J-1} z_j\Big)^{B_{J,\varepsilon}-1}$$

for suitable constants $C_{2,\varepsilon}, C_{3,\varepsilon}, A_{1,\varepsilon}, \dots, B_{1,\varepsilon}, \dots, B_{J,\varepsilon}$, it follows that

$$\left|\log\left(\frac{g_\varepsilon(\mathbf{z})}{\prod_{j=1}^{J-1} z_j^{\mu_{j,\varepsilon}-1}\left(1-\sum_{j=1}^{J-1} z_j\right)^{\mu_{J,\varepsilon}-1}}\right)\right| \le C_{4,\varepsilon}\left[1+\sum_{j=1}^{J-1}|\log(z_j)|+|\log(1-\sum_{j=1}^{J-1} z_j)|\right]$$
$$= C_{4,\varepsilon}\left[1+\left|\log\left(\prod_{j=1}^{J-1} z_j\Big(1-\sum_{j=1}^{J-1} z_j\Big)\right)\right|\right].$$

Combining all the estimates, one gets

$$\int \log\Big(\frac{h_{G_\varepsilon}(\mathbf{z})}{I_\varepsilon(\mathbf{z})}\Big)h_0(\mathbf{z})d\mathbf{z} \le C_{5,\varepsilon}\left[1+\int\left|\log\left(\prod_{j=1}^{J-1} z_j\Big(1-\sum_{j=1}^{J-1} z_j\Big)\right)\right| h_0(\mathbf{z})d\mathbf{z}\right] < +\infty$$

by assumption (A-15). Hence

$$0 < \int \log\Big(\frac{h_{G_\varepsilon}(\mathbf{z})}{\inf_{\boldsymbol{\theta}\in D_\varepsilon} h(\mathbf{z}|\boldsymbol{\theta})}\Big)h_0(\mathbf{z})d\mathbf{z} < +\infty.$$

Verification of (H3) of Theorem A.4. It follows immediately that, for every compact set C in the open simplex Δ^J,

$$\inf_{\mathbf{z}\in C}\inf_{\boldsymbol{\theta}\in D_\varepsilon} h(\mathbf{z}|\boldsymbol{\theta}) \ge \inf_{\mathbf{z}\in C} I^*_\varepsilon(\mathbf{z})$$

and the right hand side is strictly positive.

Verification of (H4) of Theorem A.4. Under the hypotheses, the function $(\boldsymbol{\theta},\mathbf{z}) \mapsto h(\mathbf{z}|\boldsymbol{\theta})$ is continuous and hence uniformly continuous on the compact set $C \times D_\varepsilon$. It follows that the family $\{\boldsymbol{\theta} \mapsto h(\mathbf{z}|\boldsymbol{\theta}) : \mathbf{z} \in C\}$ is uniformly equicontinuous on D_ε.

The proof of Theorem A.1 is analogous, it consists in an application of Theorem A.4 for $\mathcal{X}_0 = \mathcal{X}$. In the present case, everything has an extra dependence on the fixed $\boldsymbol{\xi}$ in $\mathcal{X}$. In place of $I_\varepsilon(\mathbf{z})$ one has

$$I_\varepsilon(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi}) := \inf_{\boldsymbol{\theta}\in D_\varepsilon}\frac{1}{c(\boldsymbol{\theta})}\prod_{j=1}^{J}\varrho_j(\boldsymbol{\theta})^{\xi_j}(1-\varrho_j(\boldsymbol{\theta}))^{1-\xi_j}\frac{1}{Z_{\boldsymbol{\theta}}(\boldsymbol{\xi})}\prod_{j\in\mathcal{J}^*} z_j^{\tilde{\nu}_j(\boldsymbol{\theta})-1},$$

where $\mathcal{J}^* = \{j = 1,\dots,J : \xi_j = 0\}$ and

$$Z_{\boldsymbol{\theta}}(\boldsymbol{\xi}) = \frac{\prod_{j\in\mathcal{J}^*}\Gamma(\tilde{\nu}_j(\boldsymbol{\theta}))}{\Gamma\Big(\sum_{j\in\mathcal{J}^*}\tilde{\nu}_j(\boldsymbol{\theta})\Big)}.$$

Moreover,

$$I_\varepsilon(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi}) \geq C_{1,\varepsilon}(\boldsymbol{\xi}) \prod_{j\in\mathcal{J}^*} z_j^{\mu_{j,\varepsilon}-1} =: I_\varepsilon^*(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\xi}),$$

where

$$C_{1,\varepsilon}(\boldsymbol{\xi}) = \inf_{\boldsymbol{\theta}\in D_\varepsilon} \frac{1}{c(\boldsymbol{\theta})} \prod_{j=1}^{J} \varrho_j(\boldsymbol{\theta})^{\xi_j}(1-\varrho_j(\boldsymbol{\theta}))^{1-\xi_j} Z_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\xi}),$$

and $\mu_{j,\varepsilon} := \sup\{\tilde{\nu}_j(\boldsymbol{\theta}) : \boldsymbol{\theta} \in D_\varepsilon\}$. Also in this case, $C_{1,\varepsilon}(\boldsymbol{\xi}) > 0$ and $\mu_{j,\varepsilon} > 0$, since D_ε is compact, $\nu_j(\boldsymbol{\theta})$ and $\varrho_j(\boldsymbol{\theta})$ are continuous, $0 < \varrho_j(\boldsymbol{\theta}) < 1$ and $\nu_j(\boldsymbol{\theta}) > 0$, $j = 1, \ldots, J$. Finally,

$$C_{2,\varepsilon}(\boldsymbol{\xi}) \prod_{j\in\mathcal{J}^*} z_j^{A_{j,\varepsilon}-1} \leq h_{G_\varepsilon}(\boldsymbol{\xi}, \mathbf{z}) \leq C_{3,\varepsilon}(\boldsymbol{\xi}) \prod_{j\in\mathcal{J}^*} z_j^{B_{j,\varepsilon}-1}$$

for suitable constants $C_{2,\varepsilon}(\boldsymbol{\xi}), C_{3,\varepsilon}(\boldsymbol{\xi}), A_{1,\varepsilon}, \ldots, B_{1,\varepsilon}, \ldots, B_{J,\varepsilon}$. With this minor modifications, the verification of (H1) and (H2) is exactly as in the proof of Theorem A.2. Assumption (H3) is true since

$$\inf_{\mathbf{z}_{\boldsymbol{\xi}}\in C_{\boldsymbol{\xi}}} \inf_{\boldsymbol{\theta}\in D_\varepsilon} h(\boldsymbol{\xi}|\boldsymbol{\theta}) h(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\theta}) \geq \inf_{\mathbf{z}\in C_{\boldsymbol{\xi}}} I_\varepsilon^*(\mathbf{z}|\boldsymbol{\xi})$$

and the right hand side is strictly positive by the assumptions on the $\nu_j(\boldsymbol{\theta})$s and $\varrho_j(\boldsymbol{\theta})$s. Analogously,

$$(\boldsymbol{\theta}, \mathbf{z}_{\boldsymbol{\xi}}) \mapsto h(\boldsymbol{\xi}|\boldsymbol{\theta}) h(\mathbf{z}_{\boldsymbol{\xi}}|\boldsymbol{\theta})$$

is uniformly continuous on the compact set $C_{\boldsymbol{\xi}} \times D_\varepsilon$ and hence (H4) follows. □

C.D.2 Proof of Proposition A.1

The proof of Proposition A.1 is divided in various Lemmata. For the sake of notational simplicity set

$$\mathcal{D}ir(\mathbf{z}|a_1, \ldots, a_J) = \frac{\Gamma\left(\sum_{j=1}^J a_j\right)}{\prod_{j=1}^J \Gamma(a_j)} \prod_{j=1}^{J-1} z_j^{a_j-1} \left(1 - \sum_{j=1}^{J-1} z_j\right)^{a_j-1}.$$

Note that

$$h(\mathbf{z}|\boldsymbol{\theta}) = \mathcal{D}ir(\mathbf{z}|\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}))$$

where $\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta})$ is defined in (A-24).

Lemma A.1. [Barrientos et al. (2015)] Let g_0 be a continuous density on Δ^J. Then, for every $\varepsilon > 0$ there is a density $g_\varepsilon(\mathbf{z}) = \sum_{i=1}^{M_\varepsilon} q_{i,\varepsilon} \mathcal{D}ir(\mathbf{z}|a_{i,1,\varepsilon}, \ldots, a_{i,J,\varepsilon})$ where $a_{i,j,\varepsilon} \geq 1$ for every i and j, such that

$$\|g_0 - g_\varepsilon\|_\infty \leq \varepsilon.$$

Lemma A.2. Let $a = (a_1, \ldots, a_J) \in [1, +\infty)^J$. If for any $\delta > 0$ there is $\boldsymbol{\theta}_\delta \in \Theta$ such that $\|a - \tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}_\delta)\|_\infty \leq \delta$ then for any $\varepsilon > 0$ there is $\boldsymbol{\theta}_\varepsilon \in \Theta$ such that

$$\|\mathcal{D}ir(\cdot|a_1, \ldots, a_J) - \mathcal{D}ir(\cdot|\tilde{\nu}_1(\boldsymbol{\theta}_\varepsilon), \ldots, \tilde{\nu}_J(\boldsymbol{\theta}_\varepsilon))\|_\infty \leq \varepsilon.$$

Proof. The Proof is left to the reader. □

Lemma A.3. Assume that, for every $a = (a_1, \ldots, a_j) \in [1, +\infty)^J$ and every $\delta > 0$ there is $\boldsymbol{\theta}_\delta \in \Theta$ such that $\|a - \tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}_\delta)\|_\infty \leq \delta$. Then, for every continuous density g_0 on Δ^J and for every $\varepsilon > 0$, there is a density $\tilde{g}_\varepsilon(\mathbf{z}) = \sum_{i=1}^{M_\varepsilon} q_{i,\varepsilon} \mathcal{D}ir(\mathbf{z}|\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}_{i,\varepsilon}))$ in $\mathcal{M}$ such that

$$\|g_0 - \tilde{g}_\varepsilon\|_\infty \leq \varepsilon.$$

Proof. By Lemma A.1, there is a density $g_\varepsilon(\mathbf{z}) = \sum_{i=1}^{M_\varepsilon} q_{i,\varepsilon} \mathcal{D}ir(\mathbf{z}|a_{i,1,\varepsilon}, \ldots, a_{i,J,\varepsilon})$ where $a_{i,j,\varepsilon} \geq 1$ for every i and j, such that $\|g_0 - g_\varepsilon\|_\infty \leq \varepsilon/2$. Now, by Lemma A.2, there are $\boldsymbol{\theta}_{i,\varepsilon}$ such that $\|\mathcal{D}ir(\cdot|a_{i,1,\varepsilon}, \ldots, a_{i,J,\varepsilon}) - \mathcal{D}ir(\cdot|\tilde{\nu}_1(\boldsymbol{\theta}_{i,\varepsilon}), \ldots, \tilde{\nu}_J(\boldsymbol{\theta}_{i,\varepsilon}))\|_\infty \leq \varepsilon/2$. Hence, setting $\tilde{g}_\varepsilon(\mathbf{z}) := \sum_{i=1}^{M_\varepsilon} q_{i,\varepsilon} \mathcal{D}ir(\mathbf{z}|\tilde{\nu}_1(\boldsymbol{\theta}_{i,\varepsilon}), \ldots, \tilde{\nu}_J(\boldsymbol{\theta}_{i,\varepsilon}))$, one gets

$$\begin{aligned}\|g_0 - \tilde{g}_\varepsilon\|_\infty \leq & \|g_0 - g_\varepsilon\|_\infty \\ & + \sum_{i=1}^{M} q_i \|\mathcal{D}ir(\cdot|a_{i,1,\varepsilon}, \ldots, a_{i,J,\varepsilon}) - \mathcal{D}ir(\cdot|\tilde{\nu}_1(\boldsymbol{\theta}_{i,\varepsilon}), \ldots, \tilde{\nu}_J(\boldsymbol{\theta}_{i,\varepsilon}))\|_\infty \leq \varepsilon.\end{aligned}$$

□

Lemma A.4. For every densities g_1 and g_2 in Δ^J

$$KL(g_1, g_2) \leq \frac{\sup_{\mathbf{z}} |g_1(\mathbf{z}) - g_2(\mathbf{z})|^2}{\inf_{\mathbf{z}} g_2(\mathbf{z})}$$

Proof. By Jensen inequality

$$KL(g_1, g_2) \leq \log\left(\int \frac{g_1^2}{g_2}\right).$$

Now, since $\log(1 + x) \leq x$ for every $x > 0$

$$\log\left(\int \frac{g_1^2}{g_2}\right) = \log\left(\int \left(\frac{(g_1 - g_2)^2}{g_2} + 1\right)\right) \leq \int \frac{(g_1 - g_2)^2}{g_2} \leq \frac{\sup_{\mathbf{z}} |g_1(\mathbf{z}) - g_2(\mathbf{z})|^2}{\inf_{\mathbf{z}} g_2(\mathbf{z})}$$

□

Proof of Proposition A.1. We need to prove that, if h_0 is a continuous density on Δ^J, then, for every $\eta > 0$, there is a density g_η in $\mathcal{M}$ such that

$$KL(h_0, g_\eta) \leq \eta.$$

Let $h_\varepsilon(\mathbf{z}) = \max(\varepsilon, h_0(\mathbf{z}))C_\varepsilon^{-1}$ where $C_\varepsilon := \int \max(\varepsilon, h_0(\mathbf{z}))d\mathbf{z} \leq 1+\varepsilon$. Clearly $h_\varepsilon > \varepsilon/(1+\varepsilon)$ and $h_0 \leq C_\varepsilon h_\varepsilon$. Hence, by Lemma 5.1. in Ghoshal et al. (1999), for any density g

$$KL(h_0, g) \leq (2+\varepsilon)\log(1+\varepsilon) + (1+\varepsilon)[KL(h_\varepsilon, g) + \sqrt{KL(h_\varepsilon, g)}]. \tag{A-25}$$

By Lemma A.3 there is a density $\tilde{g}_\varepsilon$ in $\mathcal{M}$ such that $\|h_\varepsilon - \tilde{g}_\varepsilon\|_\infty \leq \varepsilon/2$. From the previous inequality it follows that $\tilde{g}_\varepsilon \geq h_\varepsilon - \varepsilon/2 \geq \varepsilon/(2(1+\varepsilon))$. Hence, by Lemma A.4

$$KL(h_\varepsilon, \tilde{g}_\varepsilon) \leq \varepsilon$$

for $\varepsilon < 1$. The thesis follows by taking $\eta = (2+\varepsilon)\log(1+\varepsilon)+(1+\varepsilon)(\varepsilon+\sqrt{\varepsilon})$ and $g_\eta = \tilde{g}_\varepsilon$. □

C.D.3 Proofs of Proposition A.2 and Theorem A.3

Proof of Proposition A.2. Recall that since $Z_{i,\infty}(dy)$ is a Dirichlet process with concentration parameter $\phi(\boldsymbol{\theta}_i)$ and base measure $F(dy|\boldsymbol{\theta}_i)$, then for any finite partition $B_1, \ldots, B_J$ of $\mathbb{R}$ it follows that $(Z_{i,\infty}(B_1), \ldots, Z_{i,\infty}(B_J))$ has a Dirichlet distribution on Δ^J of parameters $(\phi(\boldsymbol{\theta}_i)F(B_1|\boldsymbol{\theta}_i), \ldots, \phi(\boldsymbol{\theta}_i)F(B_J|\boldsymbol{\theta}_i)$. Hence, the random vector $\mathbf{z}_i = (z_{i1}, \ldots, z_{iJ}) := (Z_{i,\infty}(y_1) - Z_{i,\infty}(y_0), \ldots, Z_{i,\infty}(y_J) - Z_{i,\infty}(y_{J-1}))$ has the Dirichlet distribution on the simplex Δ^J of parameters $(\phi(\boldsymbol{\theta}_i)\nu_1(\boldsymbol{\theta}_i), \ldots, \phi(\boldsymbol{\theta}_i)\nu_J(\boldsymbol{\theta}_i))$. When $\varrho_j = 0$ for $j = 1, \ldots, J$, our Bayesian model is

$$\begin{aligned}(z_{i1}, \ldots, z_{iJ}) &\sim \mathcal{D}ir(\phi(\boldsymbol{\theta}_i)\nu_1(\boldsymbol{\theta}_i), \ldots, \phi(\boldsymbol{\theta}_i)\nu_J(\boldsymbol{\theta}_i)) \\ \boldsymbol{\theta}_i &\overset{i.i.d.}{\sim} G \\ G &\sim \mathcal{DP}(\psi, G_0),\end{aligned}$$

and the thesis follows. □

Proof of Theorem A.3. The thesis is easily deduced from Proposition A.2. □

C.D.4 Proofs of Propositions A.3 and A.4

Proof of Proposition A.3. Note that

$$F_{n+1}(y) = E[F(y|\boldsymbol{\theta}^*_{d_{n+1}})|\mathbf{z}_i, i = 1, \dots, n]$$

which yields

$$\begin{aligned} E[F(y|\boldsymbol{\theta}^*_{d_{n+1}})|\mathbf{z}_i, i = 1, \dots, n] &= E[E[F(y|\boldsymbol{\theta}^*_{d_{n+1}})|\boldsymbol{\theta}^*_{d_i}, \mathbf{z}_i, i = 1 \dots, n]|\mathbf{z}_i, i = 1, \dots, n] \\ &= E[E[F(y|\boldsymbol{\theta}^*_{d_{n+1}})|\boldsymbol{\theta}^*_{d_i}, i = 1, \dots, n]|\mathbf{z}_i, i = 1, \dots, n] \end{aligned}$$

By Proposition A.2, $\boldsymbol{\theta}_i := \boldsymbol{\theta}^*_{d_i}$ are drawn form a $\mathcal{DP}(\psi, G_0)$, hence the predictive distribution of $\boldsymbol{\theta}^*_{d_{n+1}}$ given $\boldsymbol{\theta}^*_{d_i}, i = 1, \dots, n$ is

$$G_{n+1}(\cdot) = \frac{n}{n+\psi} \sum_{i=1}^{n} \delta_{\boldsymbol{\theta}^*_{d_i}}(d\boldsymbol{\theta}) + \frac{\psi}{n+\psi} G_0(\cdot).$$

Hence by the law of iterated expectations

$$\begin{aligned} E[F(y|\boldsymbol{\theta})|\boldsymbol{\theta}^*_{d_i}, i = 1 \dots, n] &= \int F(y|\boldsymbol{\theta}) G_{n+1}(d\boldsymbol{\theta}) \\ &= \frac{n}{n+\psi} \frac{1}{n} \sum_{i=1}^{n} F(y|\boldsymbol{\theta}^*_{d_i}) + \frac{\psi}{n+\psi} \int F(y|\theta) G_0(d\theta). \end{aligned}$$

Since $F(y|\boldsymbol{\theta}^*_{d_i}) = F_i(y)$, one has

$$E\Big[\frac{1}{n} \sum_{i=1}^{n} F(y|\boldsymbol{\theta}^*_{d_i})|\mathbf{z}_i, i = 1, \dots, n\Big] = \hat{F}_n(y)$$

and

$$F_{n+1}(y) := P\{Y_{n+1} \leq y|\mathbf{z}_i, i = 1 \dots, n\} = \frac{n}{n+\psi} \bar{F}(y) + \frac{\psi}{n+\psi} \int F(y|\theta) G_0(d\theta).$$

□

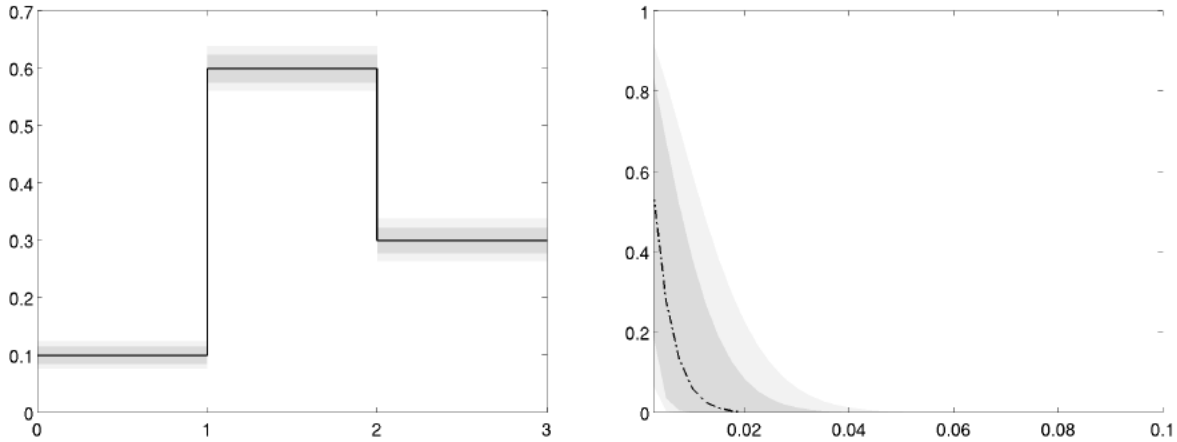
Proof of Proposition A.4. Recall that posterior consistency yields predictive consistency, see e.g. Theorem 4.2.1 in Ghosh and Ramamoorthi (2003) since $\mathbf{z} \mapsto z_i$ is a bounded and continuous function on the simplex the thesis follows. □

D Additional Results

D.A Priors

Figure A-2: Noise and zero probability

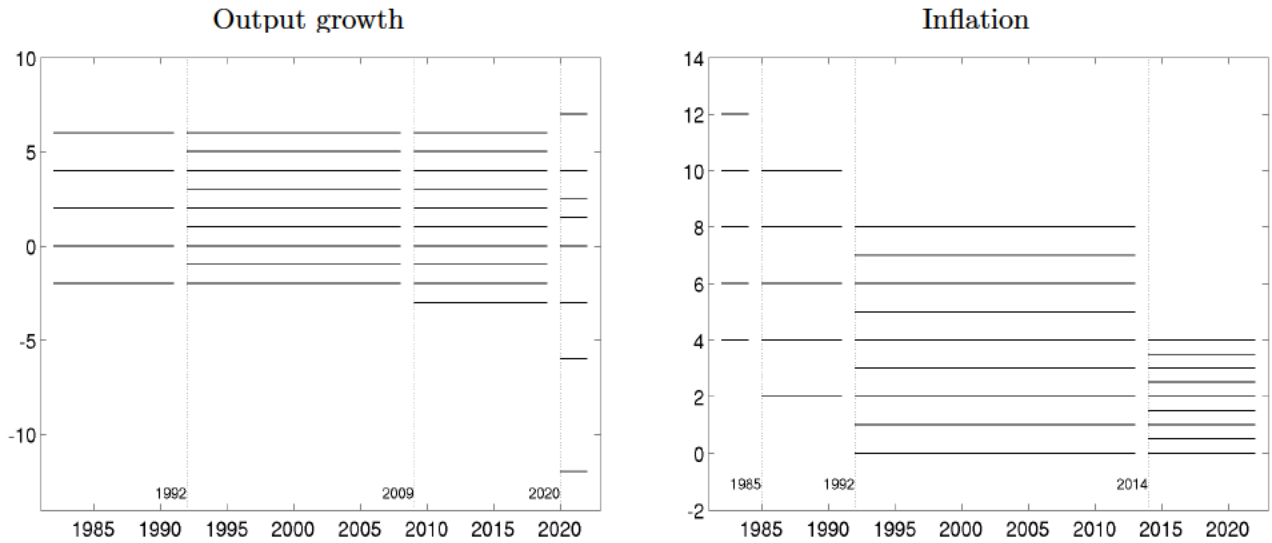
Noise around ν $\varrho(\nu,\epsilon)=\int_0^{\epsilon} b(x|\nu,r)dx$ as a function of ν



Note: The left panel of Figure A-2 shows the 50 and 90 percent a-priori coverage intervals for the noise associated with three different values of ν: 0.1, 0.6 and 0.3.

D.B Survey design

Figure A-3: Survey bins



Note: Solid lines show the survey bins. Vertical dotted lines highlight years when bins changed.

Figure A-4: Number of respondents for H1 Output growth surveys

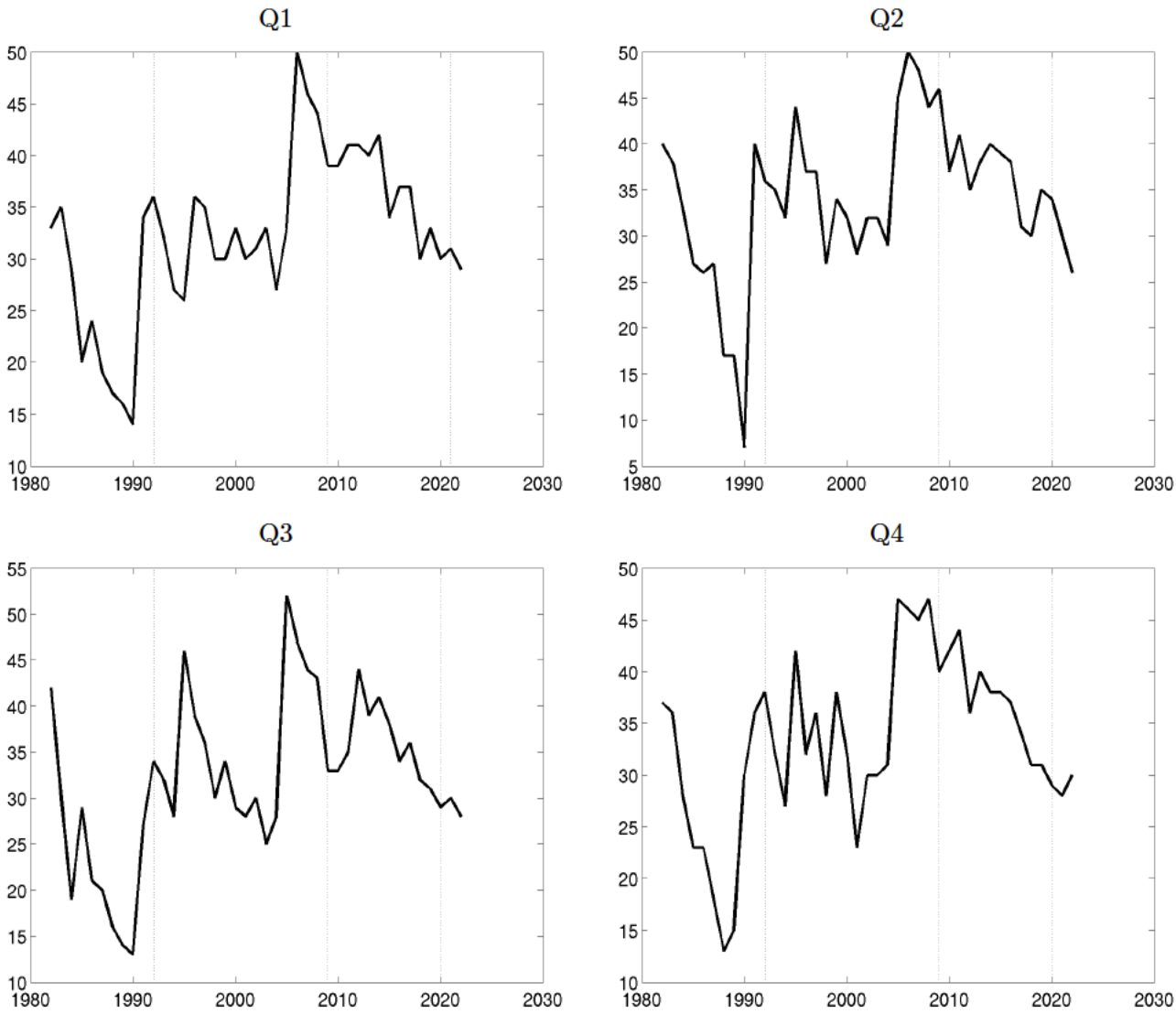
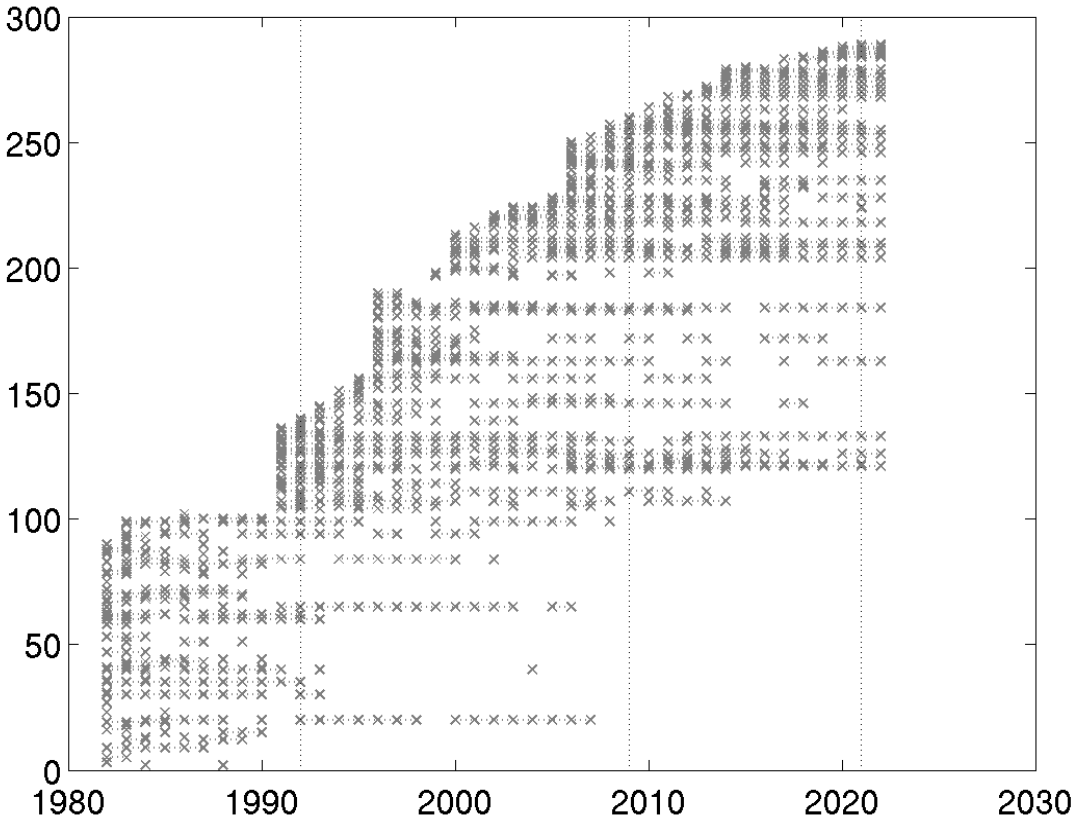


Figure A-5: SPF survey participation by respondent



Note: The light gray crosses indicate when respondents participate in a survey, and are connected by a thin dotted gray line whenever the respondent appears in consecutive surveys. Respondents are indexed by a number increasing in the year they joined the survey (y axis).

Figure A-6: Percentage of respondents for H2 Output growth surveys placing positive probability on either one open bin (solid line) or both (dash-and-dotted line)

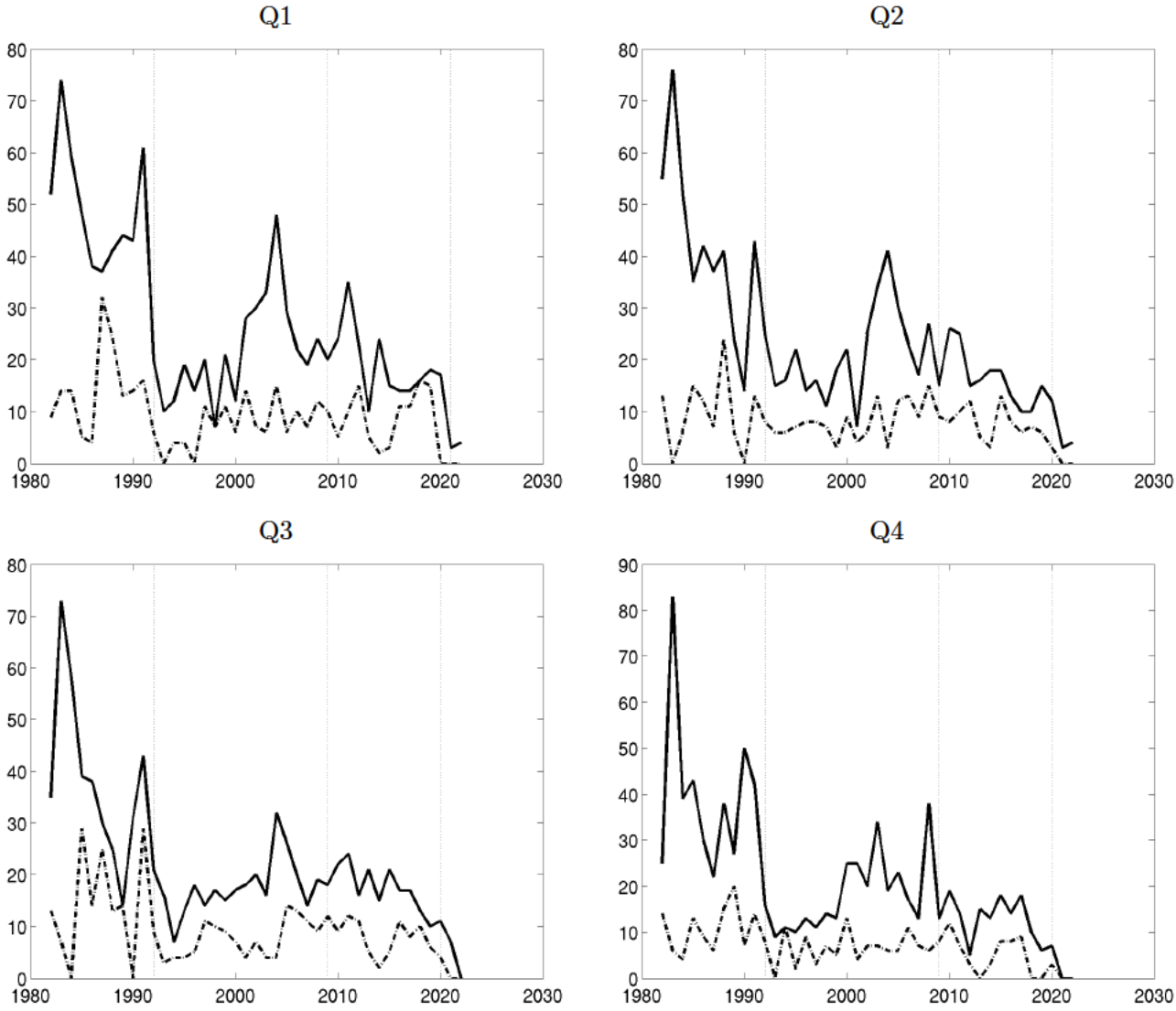
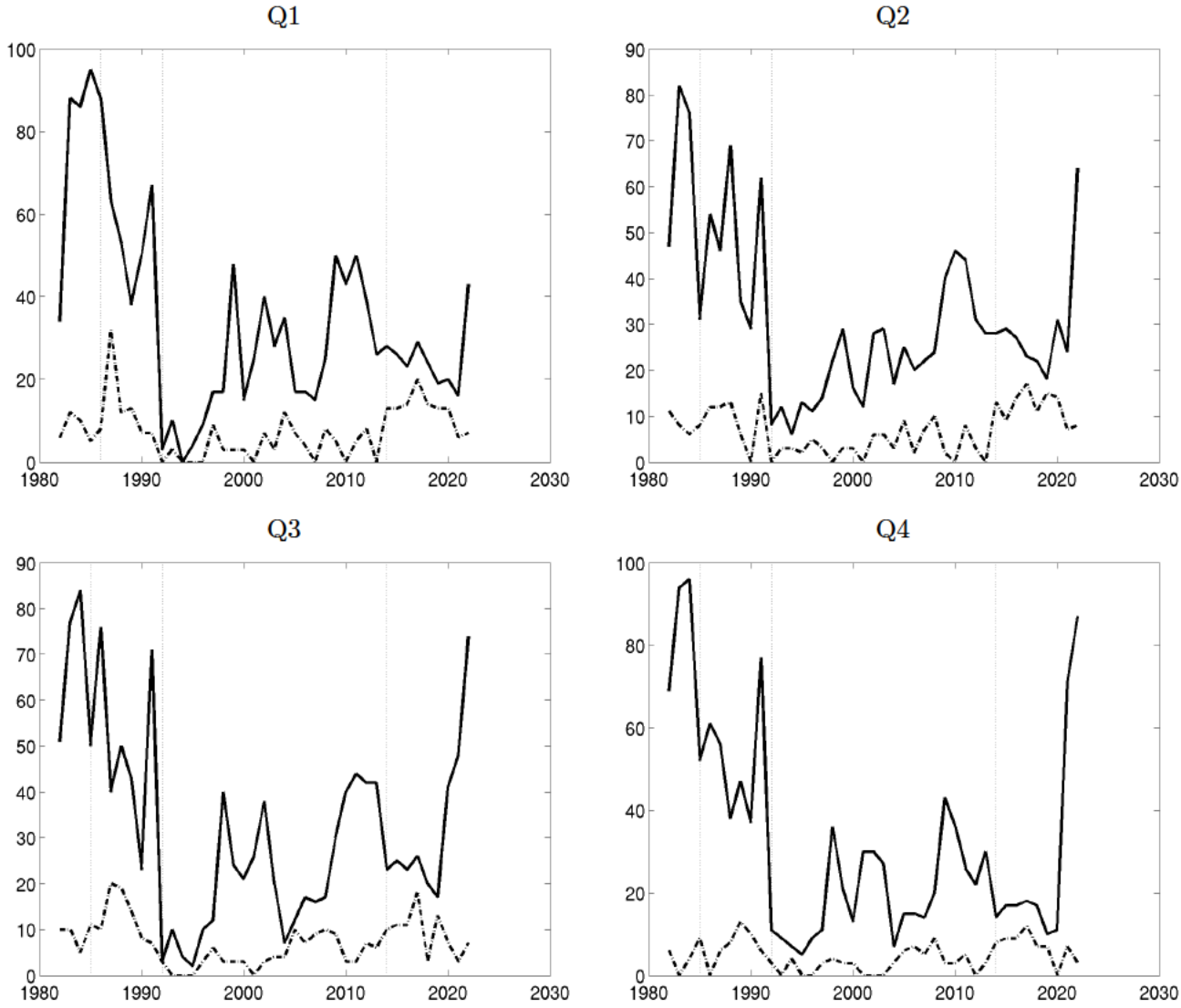
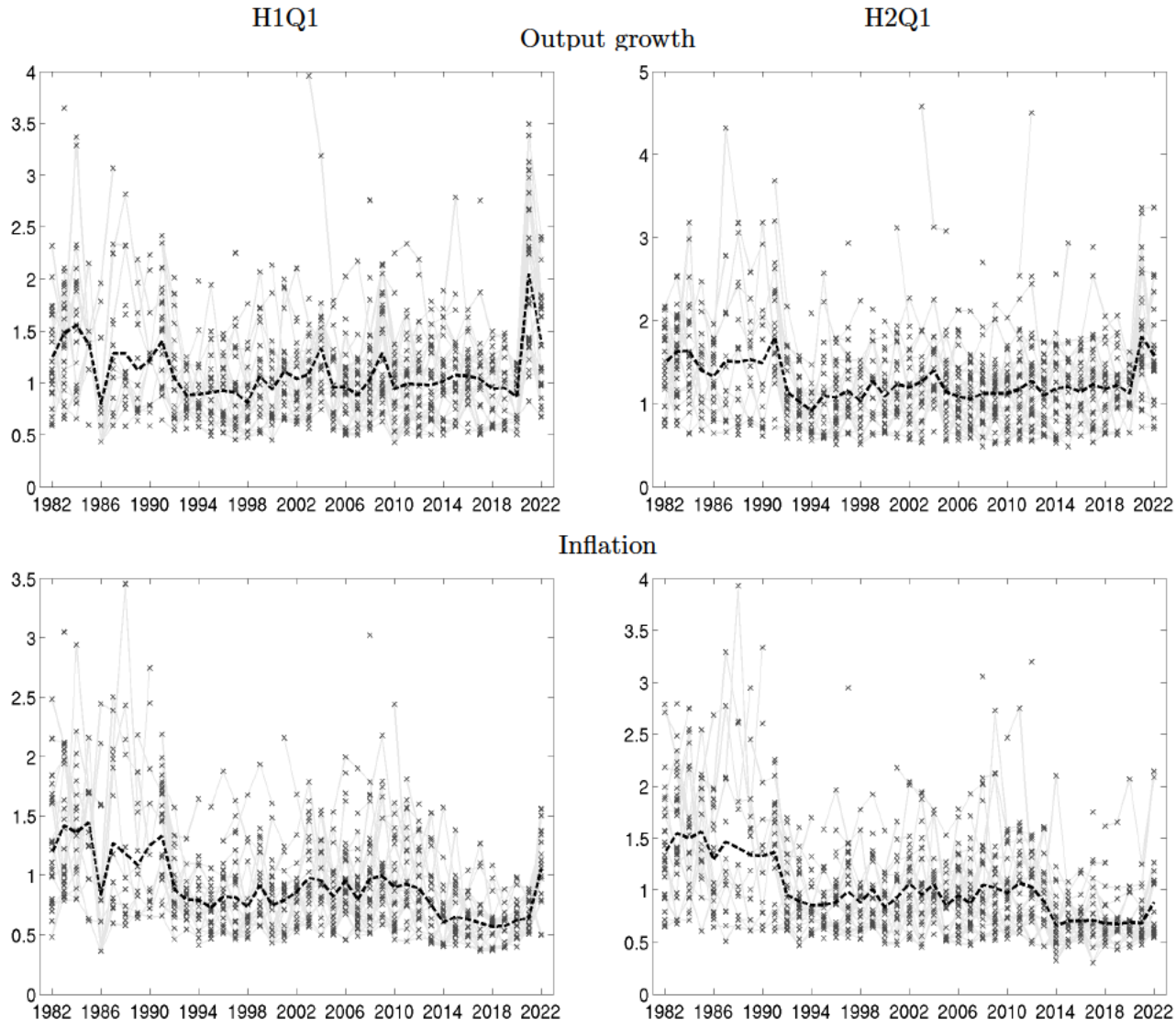


Figure A-7: Percentage of respondents for H2 inflation surveys placing positive probability on either one open bin (solid line) or both (dash-and-dotted line)



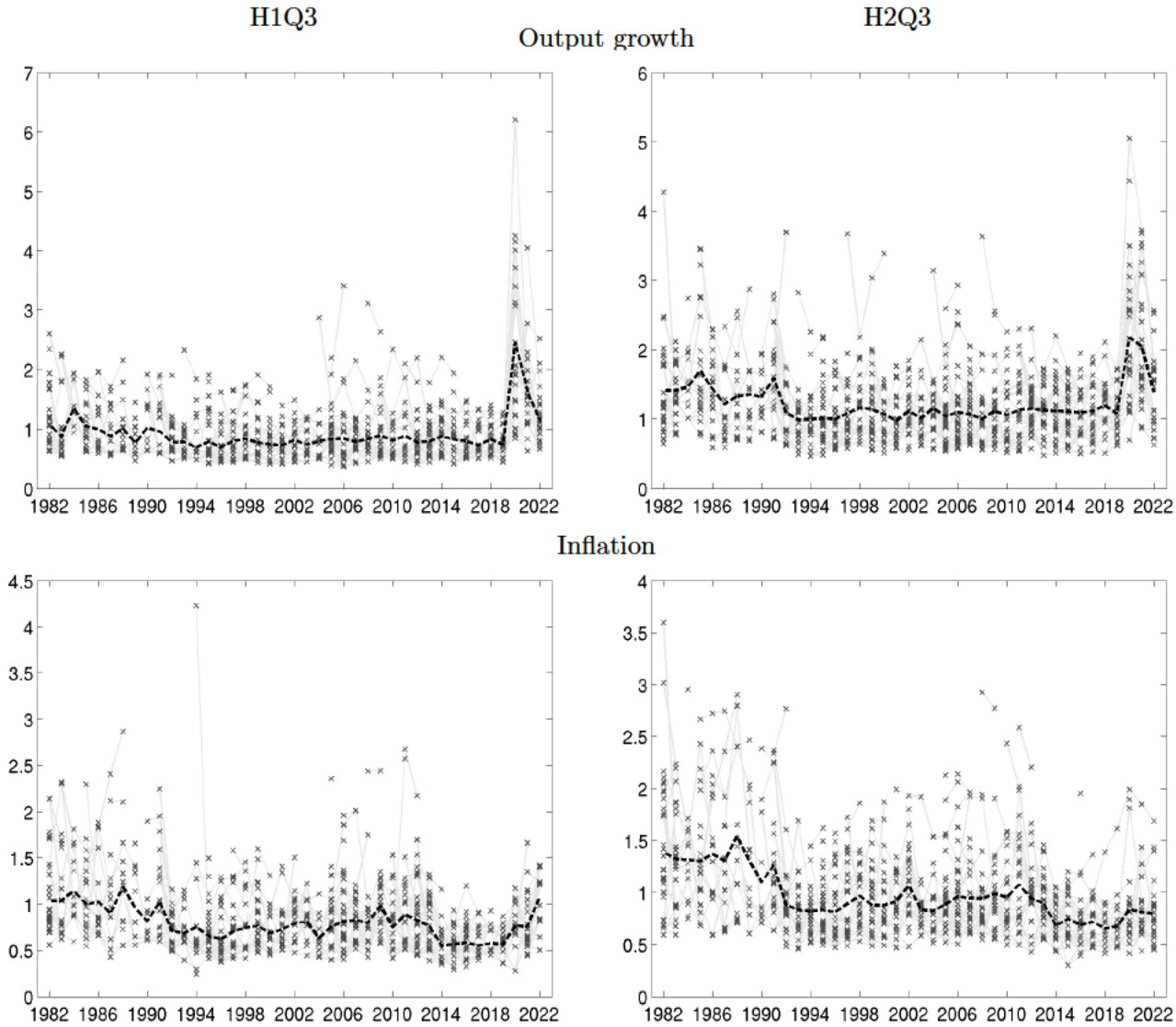
D.C Heterogeneity in subjective uncertainty

Figure A-8: Subjective uncertainty by individual respondent: Q1 survey



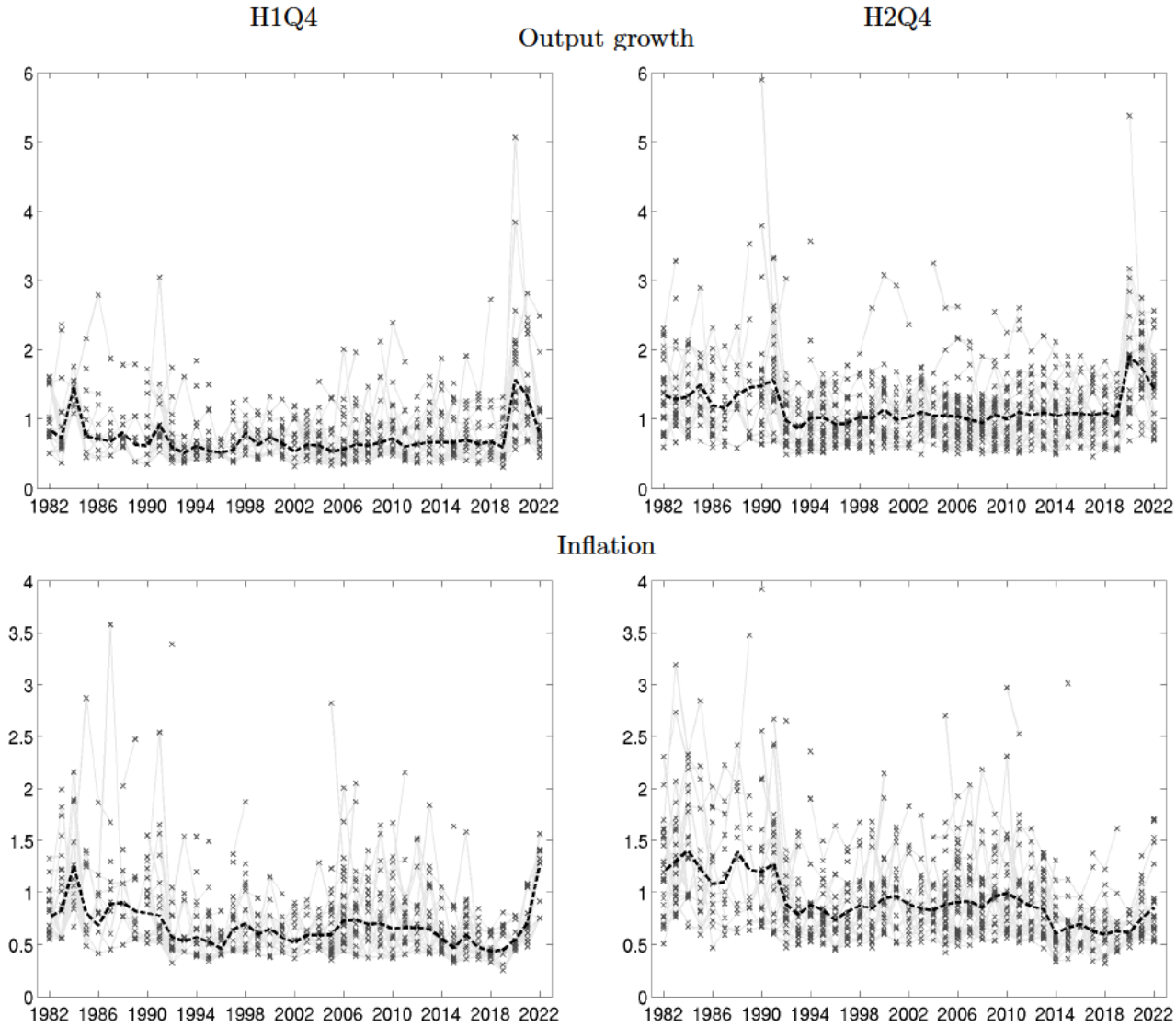
Note: Each panel displays the posterior mean of the standard deviation of the subjective predictive distribution by individual respondent (light gray crosses, connected by thin gray line whenever the respondent appears in consecutive surveys), and the cross-sectional average of the individual standard deviations (dashed black line). Top panels: Output growth projections; bottom panels: inflation projections. Left column: current year projections; left column: following year projections.

Figure A-9: Subjective uncertainty by individual respondent: Q3



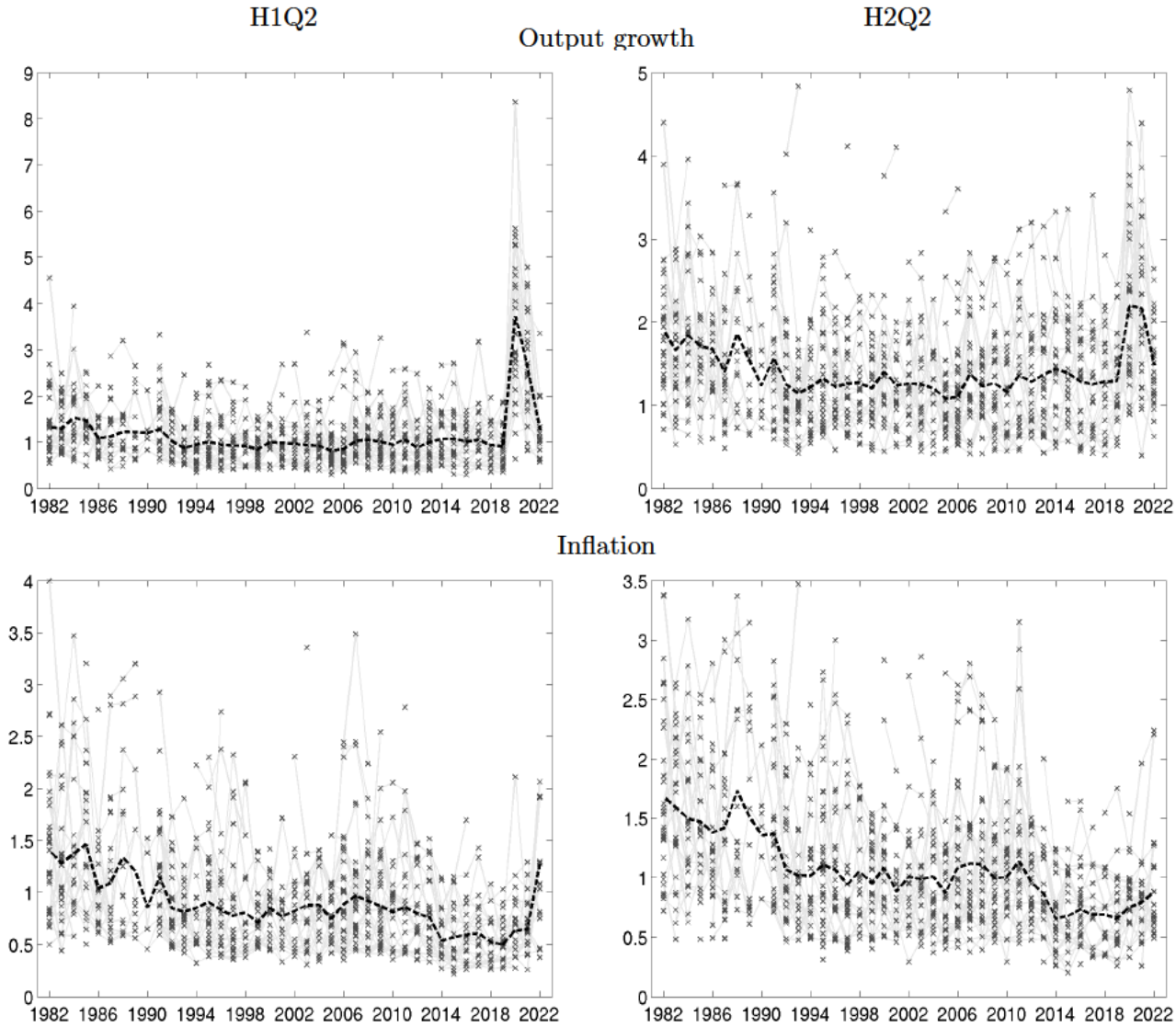
Note: Each panel displays the posterior mean of the standard deviation of the subjective predictive distribution by individual respondent (light gray crosses, connected by thin gray line whenever the respondent appears in consecutive surveys), and the cross-sectional average of the individual standard deviations (dashed black line). Top panels: Output growth projections; bottom panels: inflation projections. Left column: current year projections; left column: following year projections.

Figure A-10: Subjective uncertainty by individual respondent: Q4



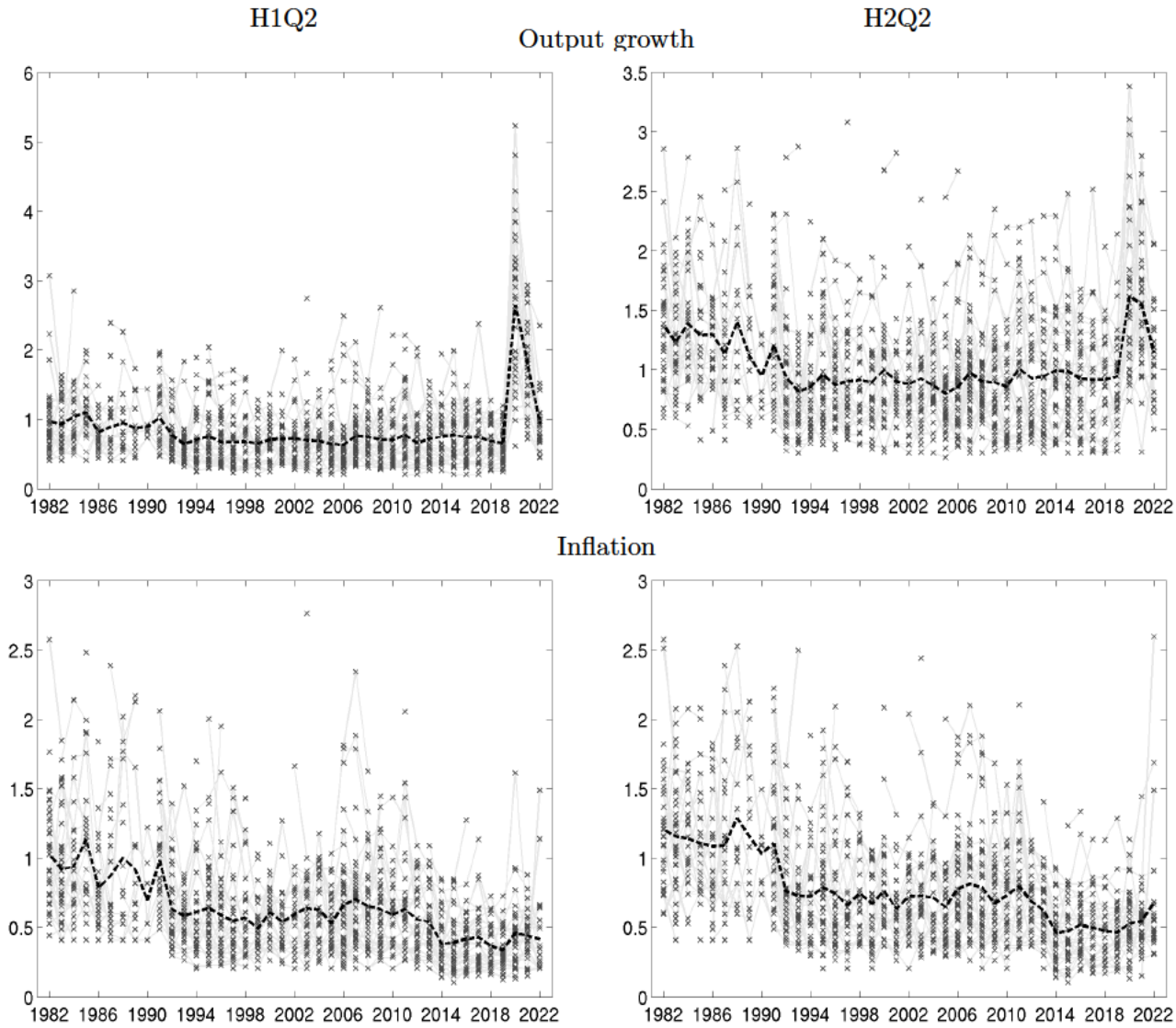
Note: Each panel displays the posterior mean of the standard deviation of the subjective predictive distribution by individual respondent (light gray crosses, connected by thin gray line whenever the respondent appears in consecutive surveys), and the cross-sectional average of the individual standard deviations (dashed black line). Top panels: Output growth projections; bottom panels: inflation projections. Left column: current year projections; left column: following year projections.

Figure A-11: Subjective uncertainty by individual respondent: IQRs, Q2 survey



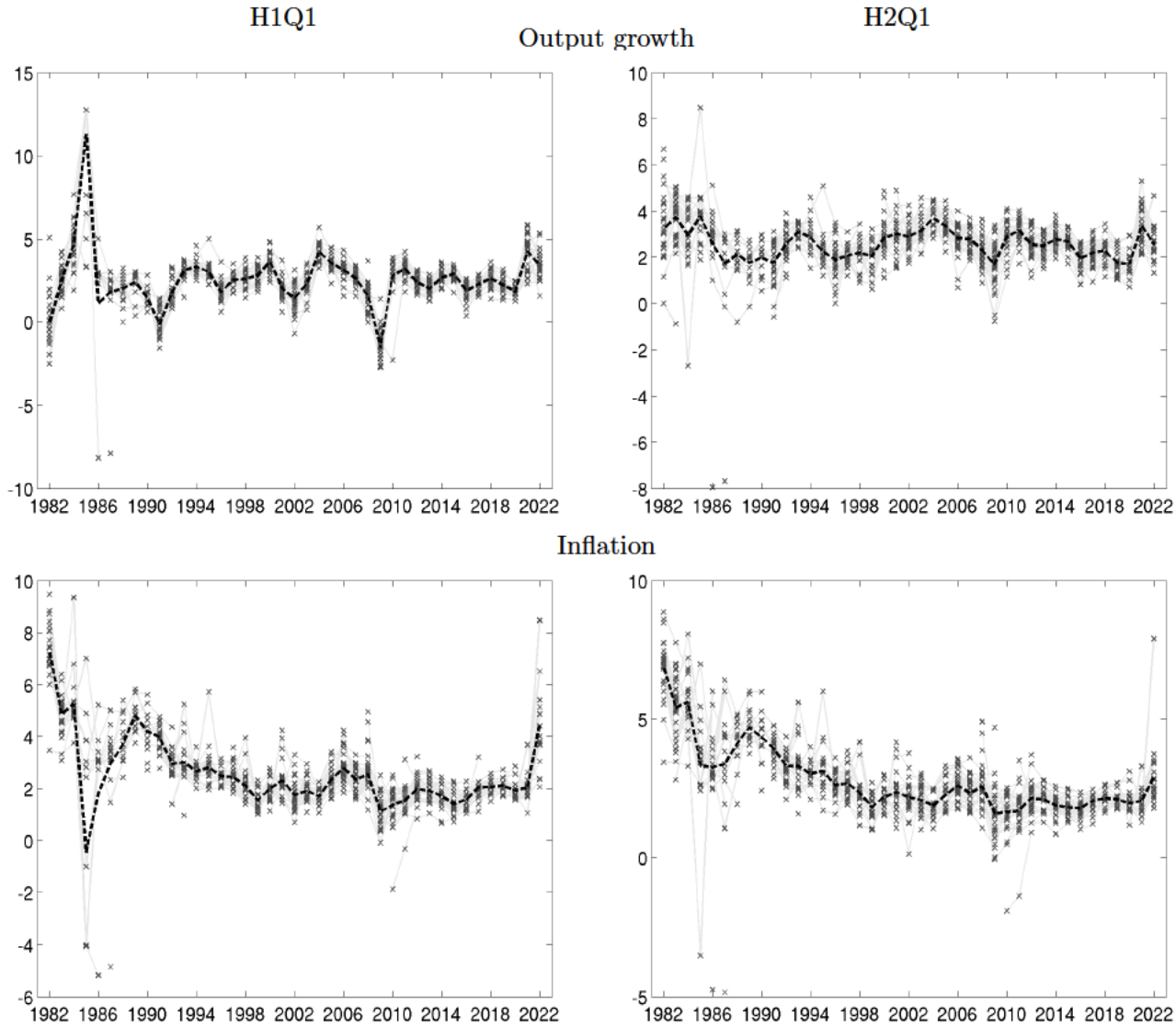
Note: Each panel displays the posterior mean of the interquartile range of the subjective predictive distribution by individual respondent (light gray crosses, connected by thin gray line whenever the respondent appears in consecutive surveys), and the cross-sectional average of individual interquartile ranges (dashed black line). Top panels: Output growth projections; bottom panels: inflation projections. Left column: current year projections; left column: following year projections.

Figure A-12: Subjective uncertainty by individual respondent: Beta, Q2 survey



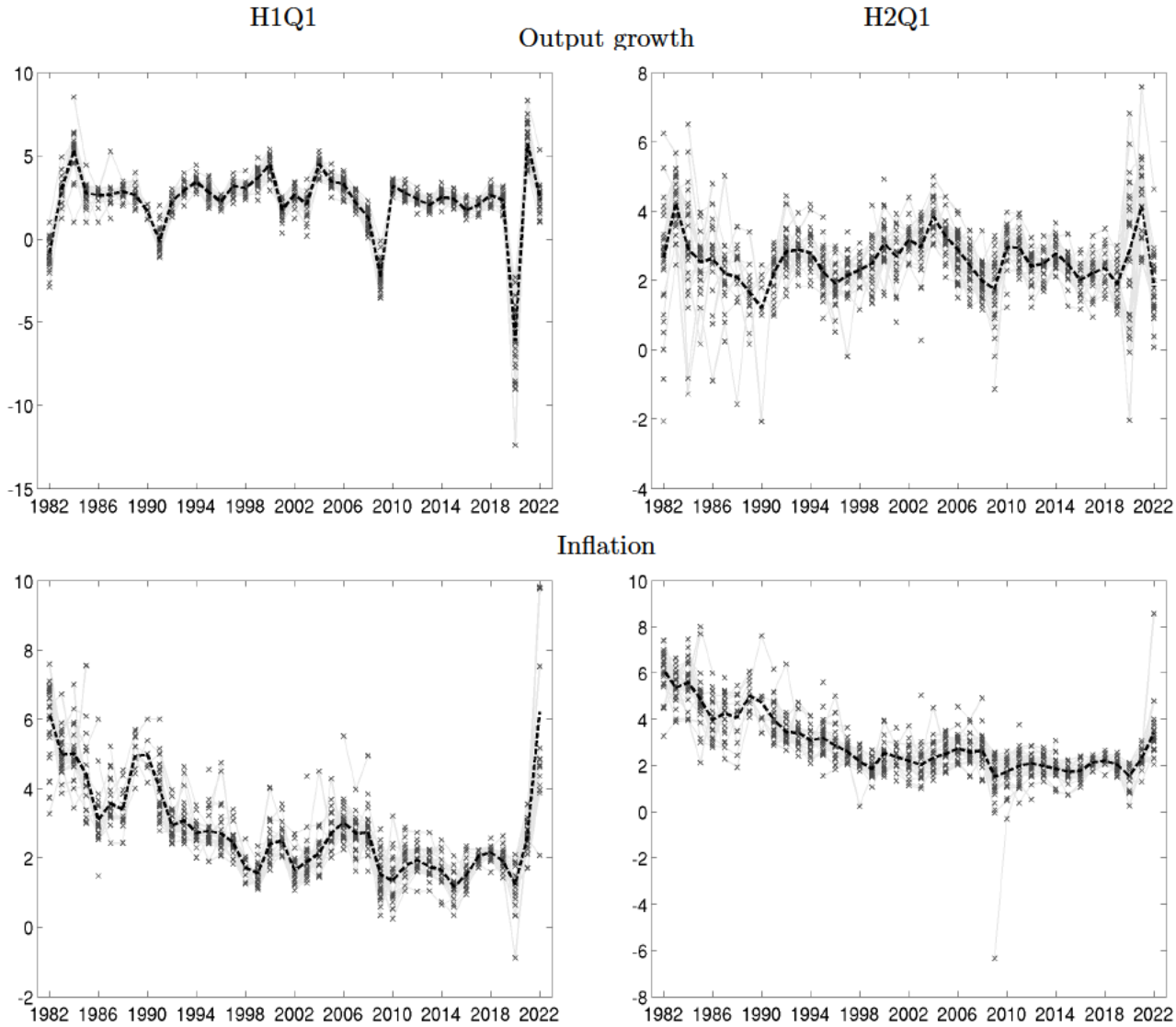
Note: Each panel displays the standard deviation of the subjective predictive distribution by individual respondent when this is estimated using least-squares under the beta parametric assumption (light gray crosses, connected by thin gray line whenever the respondent appears in consecutive surveys), and the cross-sectional average of the individual standard deviations (dashed black line). Top panels: Output growth projections; bottom panels: inflation projections. Left column: current year projections; left column: following year projections.

Figure A-13: Mean predictions by individual respondent: Q1 survey



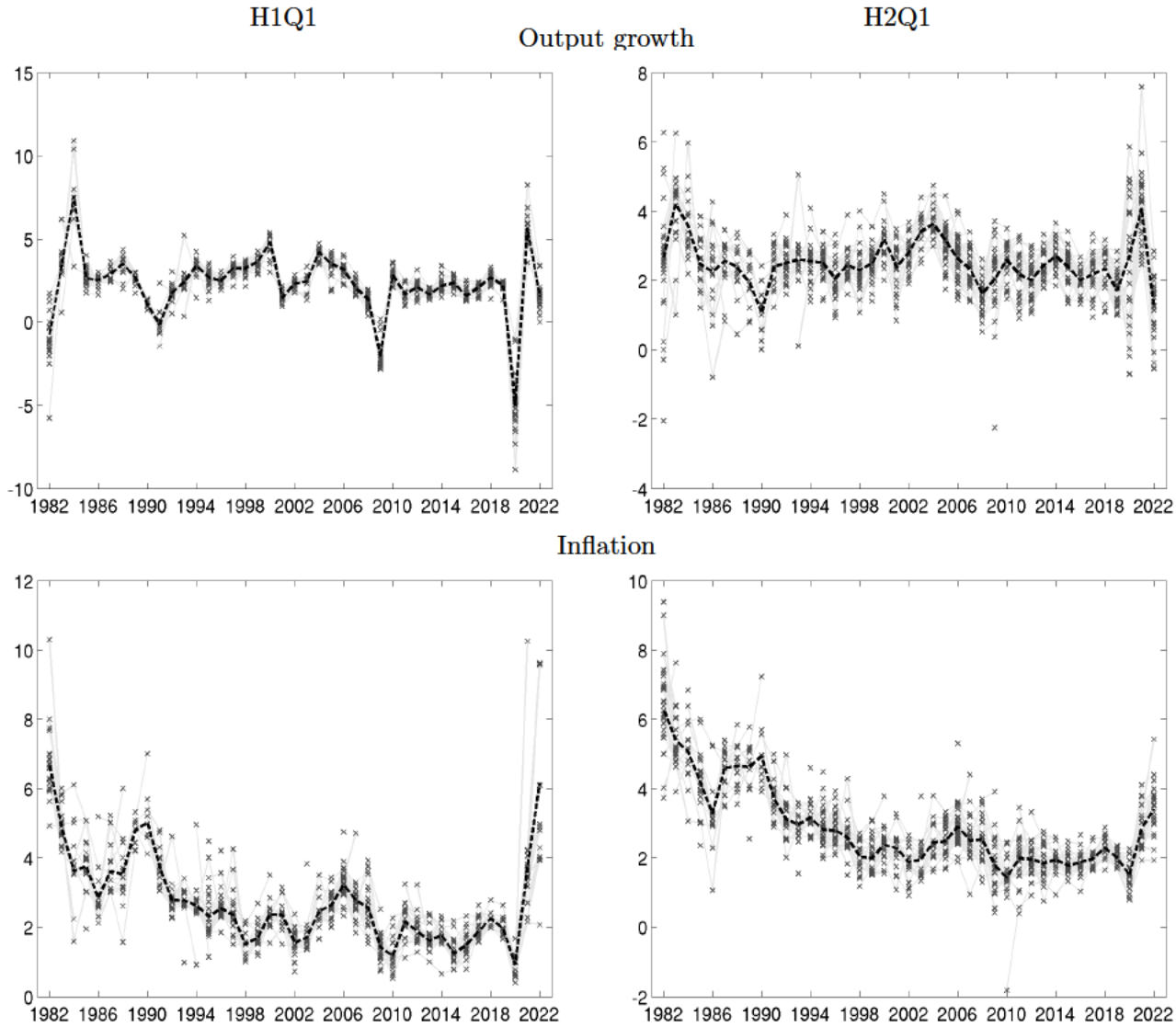
Note: Each panel displays the posterior mean of the mean of the subjective predictive distribution by individual respondent (light gray crosses, connected by thin gray line whenever the respondent appears in consecutive surveys), and the cross-sectional average of the individual means (dashed black line). Top panels: Output growth projections; bottom panels: inflation projections. Left column: current year projections; left column: following year projections.

Figure A-14: Mean predictions by individual respondent: Q2 survey



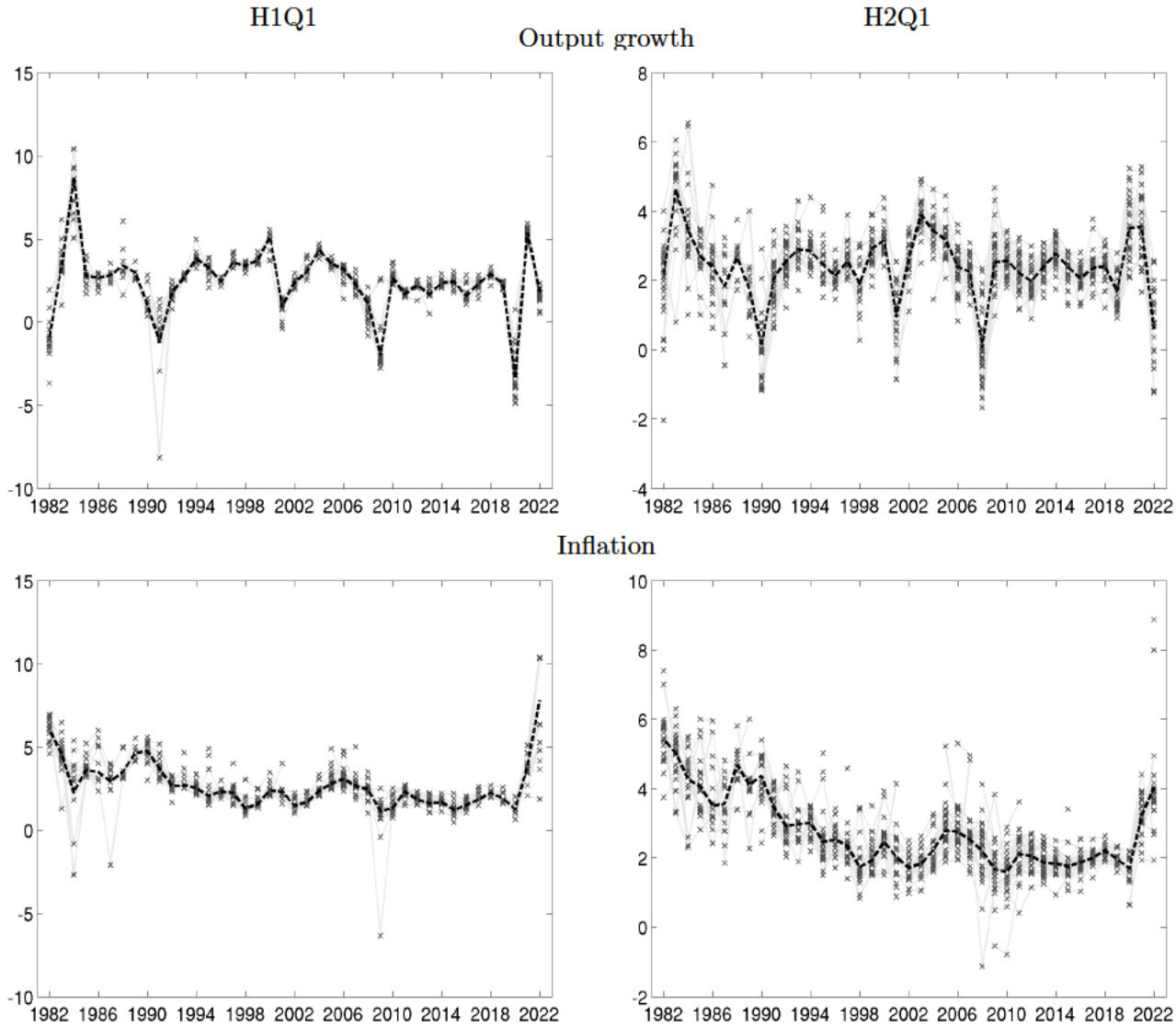
Note: Each panel displays the posterior mean of the mean of the subjective predictive distribution by individual respondent (light gray crosses, connected by thin gray line whenever the respondent appears in consecutive surveys), and the cross-sectional average of the individual means (dashed black line). Top panels: Output growth projections; bottom panels: inflation projections. Left column: current year projections; left column: following year projections.

Figure A-15: Mean predictions by individual respondent: Q3 survey



Note: Each panel displays the posterior mean of the mean of the subjective predictive distribution by individual respondent (light gray crosses, connected by thin gray line whenever the respondent appears in consecutive surveys), and the cross-sectional average of the individual means (dashed black line). Top panels: Output growth projections; bottom panels: inflation projections. Left column: current year projections; left column: following year projections.

Figure A-16: Mean predictions by individual respondent: Q4 survey



Note: Each panel displays the posterior mean of the mean of the subjective predictive distribution by individual respondent (light gray crosses, connected by thin gray line whenever the respondent appears in consecutive surveys), and the cross-sectional average of the individual means (dashed black line). Top panels: Output growth projections; bottom panels: inflation projections. Left column: current year projections; left column: following year projections.

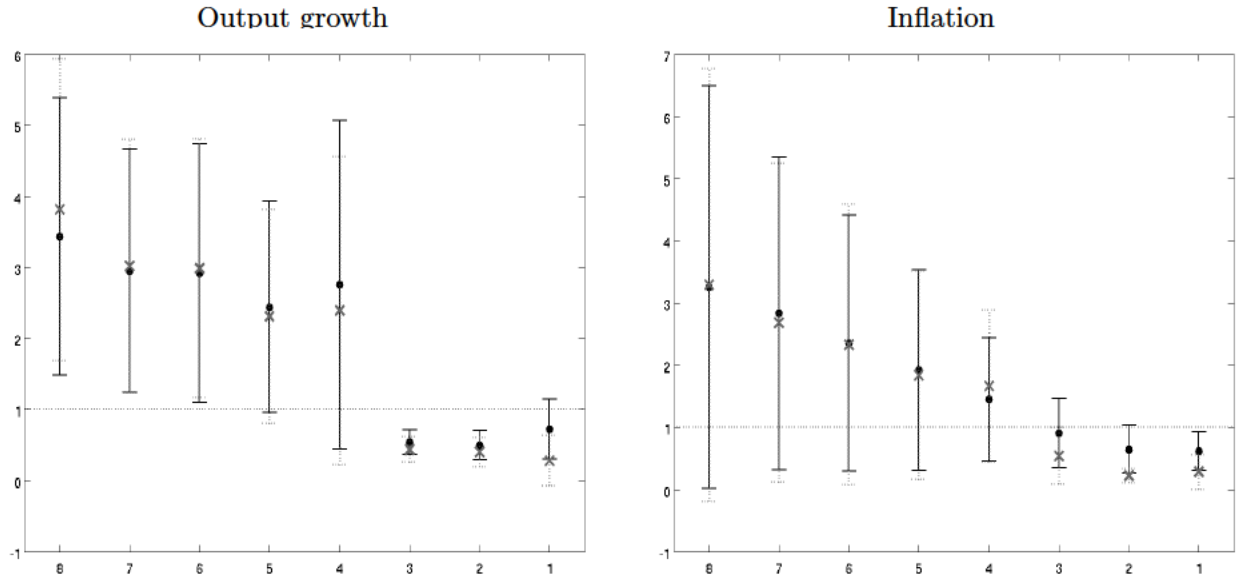
D.D Subjective uncertainty and forecast accuracy: additional results

A scale test: additional results

Table A-1: Statistics underlying Figure 5

quarters ahead:	8	7	6	5	4	3	2	1
Period:	1982- 2022							
				Output growth				
Coef.:	3.436	2.952	2.919	2.446	2.755	0.545	0.498	0.724
SE:	0.972	0.857	0.913	0.744	1.156	0.086	0.101	0.212
nobs:	1169	1271	1237	1294	1287	1371	1280	1329
no fcters:	588	588	588	589	588	591	595	592
no years:	40	40	40	40	41	41	41	41
				Inflation				
Coef.:	3.255	2.833	2.351	1.919	1.450	0.905	0.646	0.616
SE:	1.617	1.258	1.027	0.806	0.495	0.280	0.193	0.156
nobs:	1133	1235	1196	1254	1240	1316	1243	1289
no fcters:	588	588	588	589	588	591	595	592
no years:	40	40	40	40	41	41	41	41

Figure A-17: A scale test: mean vs point predictions

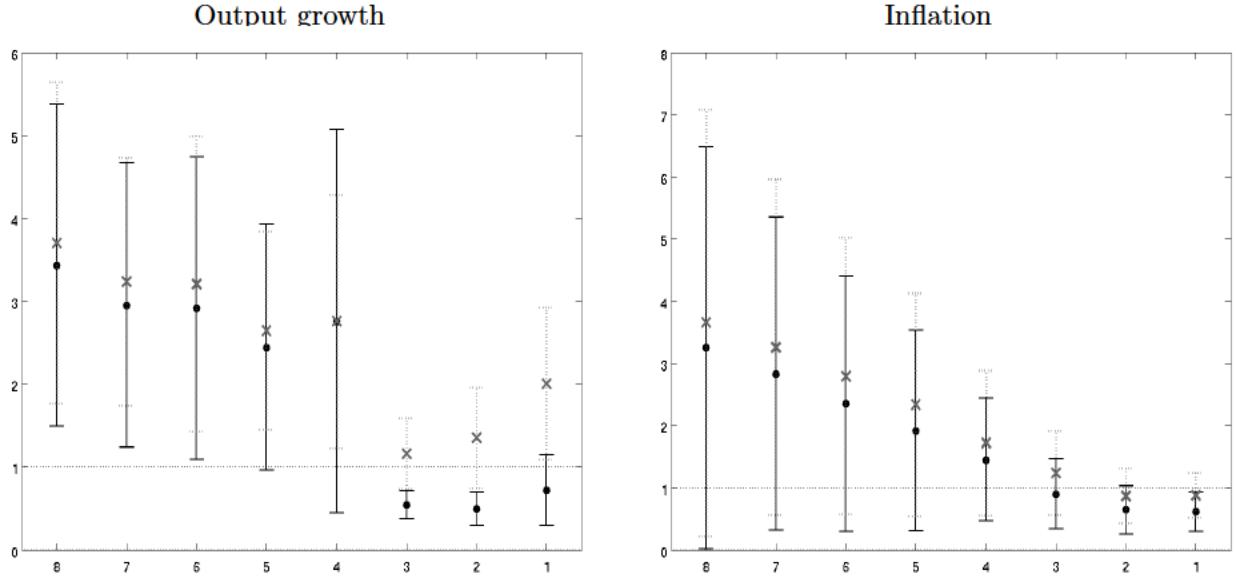


Note: Black dots correspond to OLS estimates of α_q from regression (18) for $q = 8, ..., 1$. Gray crosses correspond to OLS estimates when the point prediction $y^{pp}_{t-q,i}$ is used in place of $\mathbb{E}_{t-q,i}[y_t]$. Whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Table A-2: Statistics underlying Figure 6

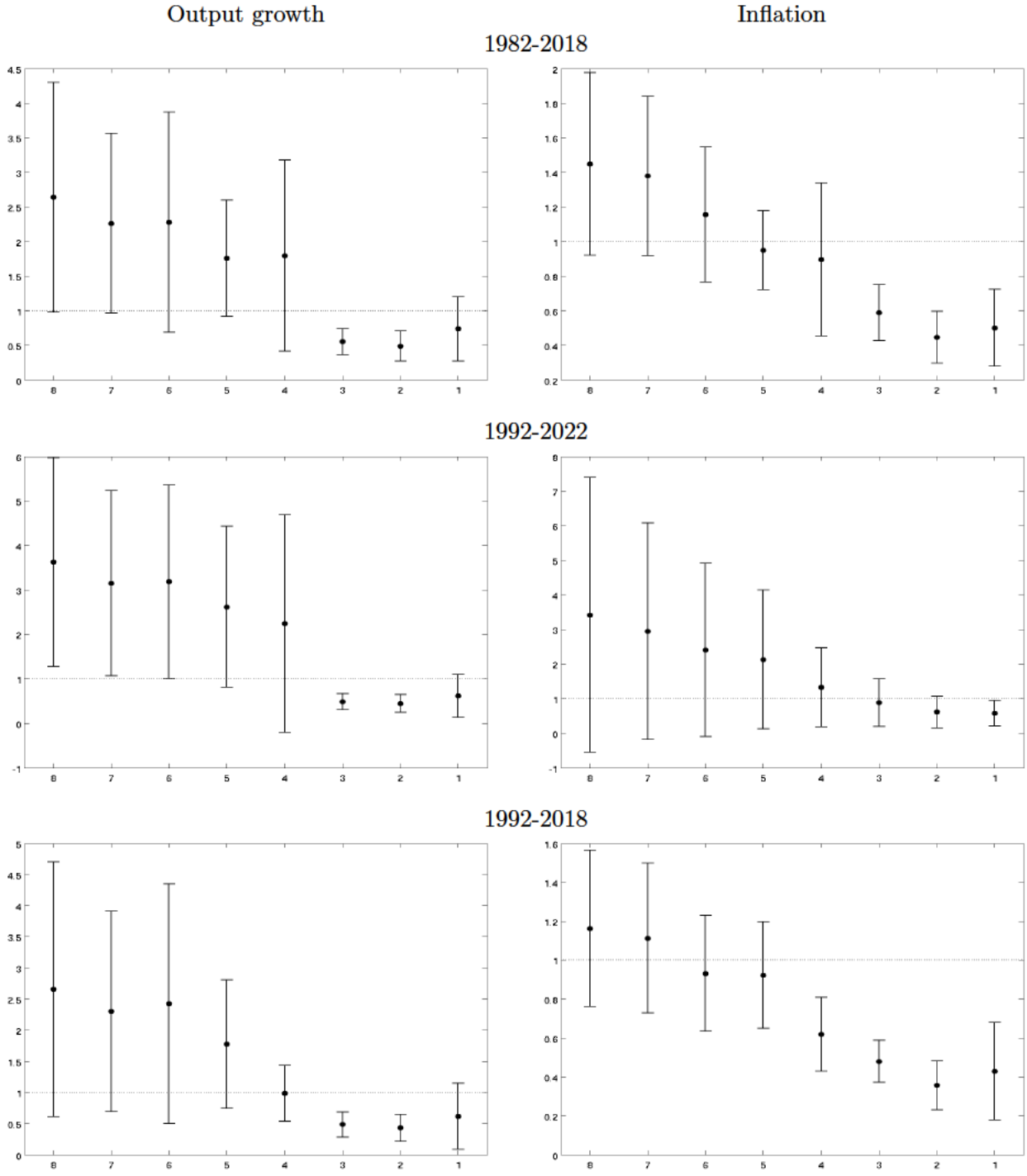
quarters ahead:	8	7	6	5	4	3	2	1
Period:	1982- 2022							
				Output growth				
Baseline								
Coef.:	3.436	2.952	2.919	2.446	2.755	0.545	0.498	0.724
SE:	0.972	0.857	0.913	0.744	1.156	0.086	0.101	0.212
nobs:	1169	1271	1237	1294	1287	1371	1280	1329
no fcters:	588	588	588	589	588	591	595	592
no years:	40	40	40	40	41	41	41	41
Beta								
Coef.:	6.577	5.591	6.102	5.014	5.043	1.170	1.110	2.098
SE:	1.947	1.706	1.967	1.606	1.879	0.189	0.228	0.655
nobs:	1169	1271	1237	1294	1287	1371	1280	1329
no fcters:	588	588	588	589	588	591	595	592
no years:	40	40	40	40	41	41	41	41
				Inflation				
Baseline								
Coef.:	3.255	2.833	2.351	1.919	1.450	0.905	0.646	0.616
SE:	1.617	1.258	1.027	0.806	0.495	0.280	0.193	0.156
nobs:	1133	1235	1196	1254	1240	1316	1243	1289
no fcters:	588	588	588	589	588	591	595	592
no years:	40	40	40	40	41	41	41	41
Beta								
Coef.:	7.652	7.711	6.206	5.675	4.590	3.629	3.460	4.366
SE:	4.102	3.817	3.146	2.795	2.261	2.008	2.108	2.885
nobs:	1133	1235	1196	1254	1240	1316	1243	1289
no fcters:	588	588	588	589	588	591	595	592
no years:	40	40	40	40	41	41	41	41

Figure A-18: A scale test: first finals vs latest vintage



Note: Black dots correspond to OLS estimates of α_q from regression (18) for $q = 8, .., 1$ using first finals as a measure of actual realizations y_t. Gray crosses correspond to OLS estimates when the latest vintage is used instead. Whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-19: A scale test: different samples



Note: Black dots correspond to OLS estimates of α_q from regression (18) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

A variation test: additional results

Table A-3: Statistics underlying Figures 7 and 8

quarters ahead:	8	7	6	5	4	3	2	1
Period:	1982- 2022							
	Output growth							
	No Fixed Effects							
Coef.:	0.068	0.136	0.242	0.250	0.449	0.605	0.783	0.935
SE:	0.136	0.175	0.172	0.165	0.150	0.220	0.189	0.198
nobs:	1169	1271	1237	1294	1287	1371	1280	1329
no fcters:	588	588	588	589	588	591	595	592
no years:	40	40	40	40	41	41	41	41
R2:	0.000	0.000	0.010	0.010	0.020	0.050	0.070	0.100
	Time Fixed Effects							
Coef.:	0.039	0.004	0.115	0.277	0.247	0.212	0.582	0.562
SE:	0.087	0.088	0.082	0.100	0.073	0.108	0.126	0.141
nobs:	1169	1271	1237	1294	1287	1371	1280	1329
no fcters:	588	588	588	589	588	591	595	592
no years:	40	40	40	40	41	41	41	41
R2:	0.540	0.520	0.500	0.500	0.450	0.320	0.340	0.440
	Forecaster Fixed Effects							
Coef.:	0.175	0.210	0.261	0.165	0.265	0.926	0.893	1.148
SE:	0.224	0.282	0.275	0.221	0.193	0.244	0.362	0.165
nobs:	718	736	717	822	817	847	771	867
no fcters:	45	45	45	51	50	51	48	54
no years:	40	40	40	40	41	41	41	41
R2:	0.050	0.050	0.040	0.070	0.060	0.130	0.150	0.220
	Time and Forecaster Fixed Effects							
Coef.:	0.384	-0.075	-0.058	0.193	0.095	0.362	0.517	0.872
SE:	0.173	0.148	0.170	0.129	0.136	0.162	0.191	0.163
nobs:	718	736	717	822	817	847	771	867
no fcters:	45	45	45	51	50	51	48	54
no years:	40	40	40	40	41	41	41	41
R2:	0.610	0.580	0.540	0.550	0.490	0.380	0.420	0.470

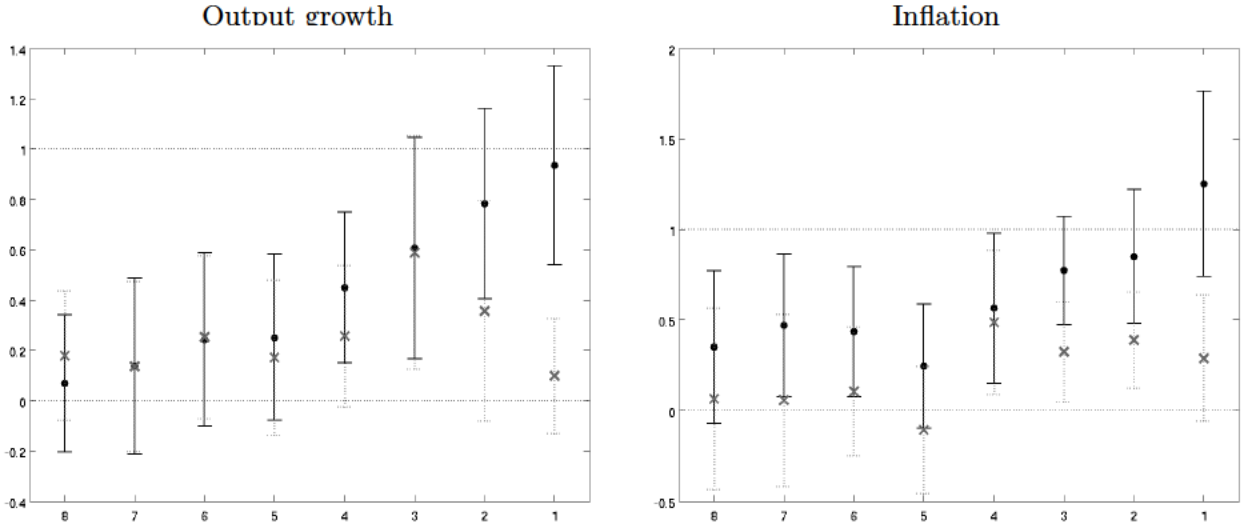
Table A-4: Table A-3 continued —Statistics underlying Figure 7

quarters ahead:	8	7	6	5	4	3	2	1
Period:	1982- 2022							
	Inflation							
	No Fixed Effects							
Coef.:	0.350	0.470	0.435	0.244	0.565	0.773	0.850	1.251
SE:	0.211	0.197	0.180	0.171	0.207	0.149	0.185	0.256
nobs:	1133	1235	1196	1254	1240	1316	1243	1289
no fcters:	588	588	588	589	588	591	595	592
no years:	40	40	40	40	41	41	41	41
R2:	0.020	0.030	0.020	0.010	0.040	0.070	0.070	0.150
	Time Fixed Effects							
Coef.:	0.244	0.302	0.264	0.194	0.416	0.505	0.715	0.957
SE:	0.086	0.103	0.103	0.107	0.140	0.130	0.165	0.133
nobs:	1133	1235	1196	1254	1240	1316	1243	1289
no fcters:	588	588	588	589	588	591	595	592
no years:	40	40	40	40	41	41	41	41
R2:	0.430	0.480	0.400	0.390	0.320	0.290	0.250	0.340
	Forecaster Fixed Effects							
Coef.:	-0.013	0.194	0.096	0.035	0.369	0.684	0.973	1.254
SE:	0.324	0.286	0.289	0.166	0.199	0.240	0.283	0.347
nobs:	705	715	662	777	804	821	746	826
no fcters:	45	44	41	48	50	50	47	51
no years:	40	40	40	40	41	41	41	41
R2:	0.050	0.080	0.080	0.070	0.090	0.140	0.160	0.200
	Time and Forecaster Fixed Effects							
Coef.:	0.224	0.229	0.008	0.024	0.298	0.332	0.831	0.930
SE:	0.139	0.136	0.160	0.168	0.220	0.172	0.233	0.168
nobs:	705	715	662	777	804	821	746	826
no fcters:	45	44	41	48	50	50	47	51
no years:	40	40	40	40	41	41	41	41
R2:	0.470	0.500	0.480	0.450	0.340	0.360	0.340	0.410

Table A-5: Statistics underlying Figure 9

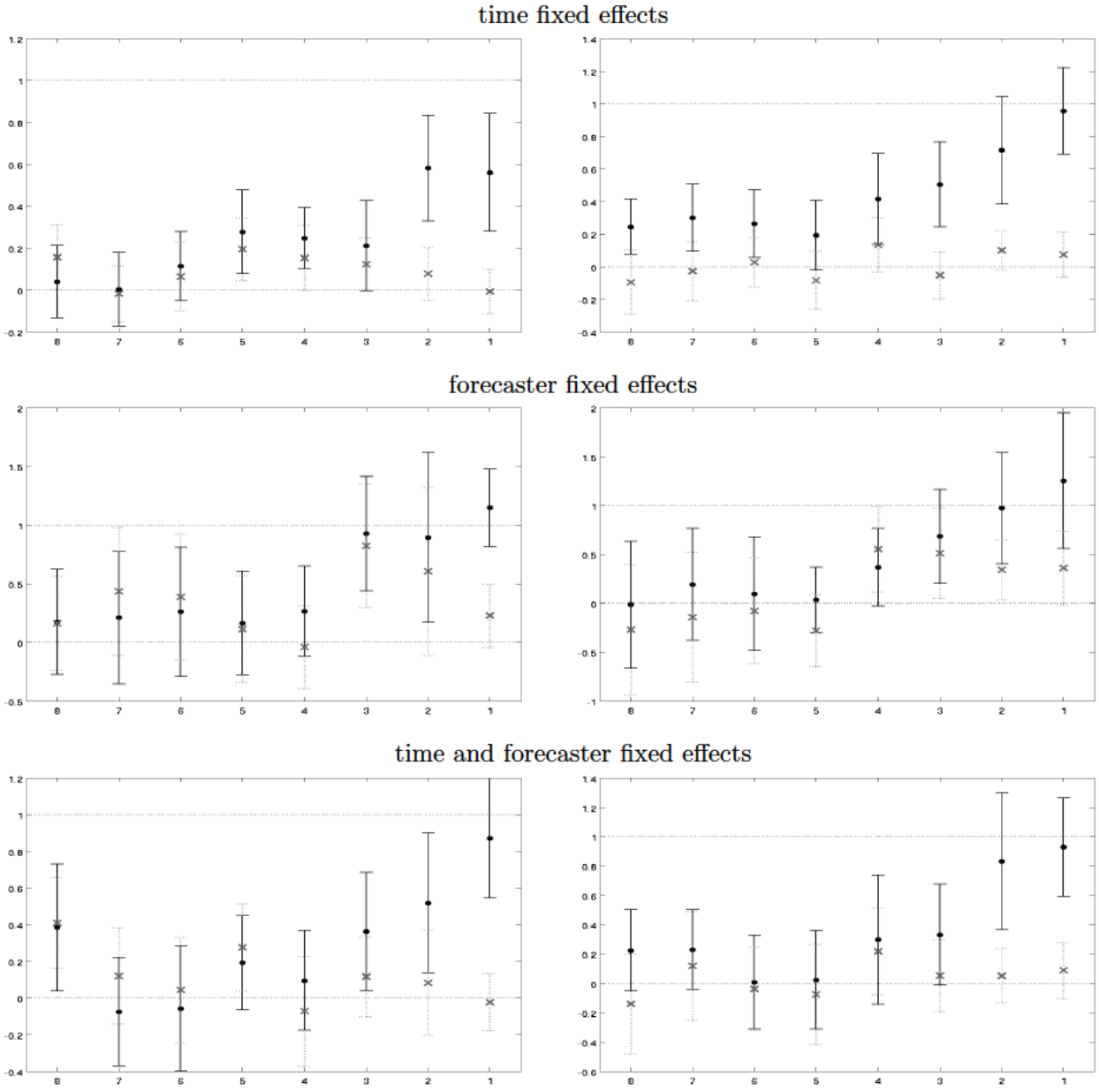
quarters ahead:	8	7	6	5	4	3	2	1
Period:	1982- 2022							
				Output growth				
Baseline								
Coef.:	0.068	0.136	0.242	0.250	0.449	0.605	0.783	0.935
SE:	0.136	0.175	0.172	0.165	0.150	0.220	0.189	0.198
nobs:	1169	1271	1237	1294	1287	1371	1280	1329
no fcters:	588	588	588	589	588	591	595	592
no years:	40	40	40	40	41	41	41	41
R2:	0.000	0.000	0.010	0.010	0.020	0.050	0.070	0.100
Weighted								
Coef.:	0.007	0.164	0.096	0.153	0.341	0.600	0.643	0.856
SE:	0.178	0.177	0.253	0.182	0.172	0.194	0.177	0.219
nobs:	1169	1271	1237	1294	1287	1371	1280	1329
no fcters:	588	588	588	589	588	591	595	592
no years:	40	40	40	40	41	41	41	41
R2:	0.000	0.000	0.000	0.000	0.010	0.030	0.040	0.070
				Inflation				
Baseline								
Coef.:	0.350	0.470	0.435	0.244	0.565	0.773	0.850	1.251
SE:	0.211	0.197	0.180	0.171	0.207	0.149	0.185	0.256
nobs:	1133	1235	1196	1254	1240	1316	1243	1289
no fcters:	588	588	588	589	588	591	595	592
no years:	40	40	40	40	41	41	41	41
R2:	0.020	0.030	0.020	0.010	0.040	0.070	0.070	0.150
Weighted								
Coef.:	0.144	0.474	0.333	0.149	0.317	0.724	0.921	1.270
SE:	0.353	0.324	0.269	0.207	0.183	0.192	0.266	0.260
nobs:	1133	1235	1196	1254	1240	1316	1243	1289
no fcters:	588	588	588	589	588	591	595	592
no years:	40	40	40	40	41	41	41	41
R2:	0.000	0.030	0.010	0.000	0.010	0.060	0.070	0.150

Figure A-20: A variation test—mean vs point projections



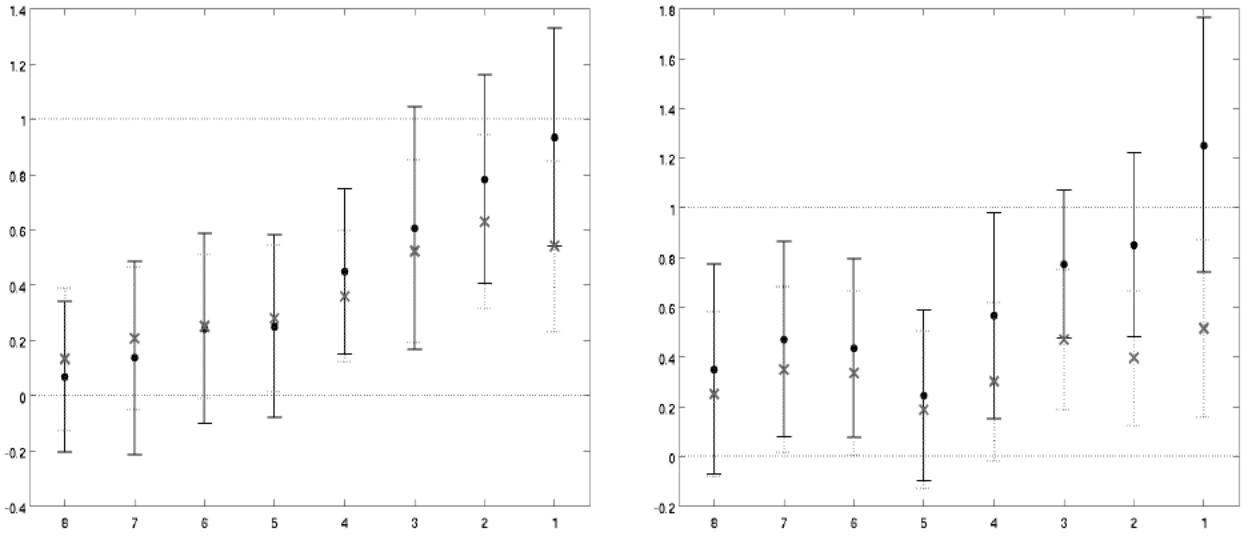
Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$. Gray crosses correspond to OLS estimates when the point prediction $y^{pp}_{t-q,i}$ is used in place of $\mathbb{E}_{t-q,i}[y_t]$. Whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-21: A variation test: Regressions with fixed effects—mean vs point projections



Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) using time (top panels), forecaster fixed effects (middle panels), or both (bottom panels), for $q = 8, ..., 1$. Gray crosses correspond to OLS estimates when the point prediction $y^{pp}_{t-q,i}$ is used in place of $\mathbb{E}_{t-q,i}[y_t]$. Dotted gray whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

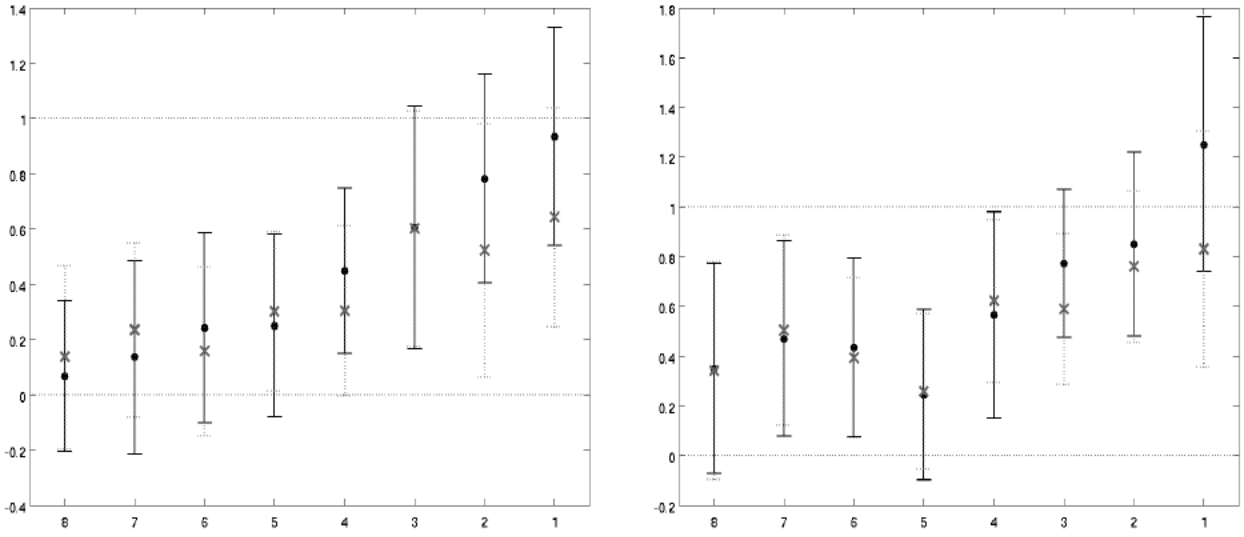
Figure A-22: A variation test: comparison with the generalized beta approach



Note: Black dots correspond to OLS estimates of β_q from regression (19) for $q = 8, ..., 1$ using the posterior means for $\mathbb{E}_{t-q,i}[y_t]$ and $\sigma_{t|t-q,i}$ from the approach in this paper. Gray crosses correspond to OLS estimates when these objects are obtained using the generalized beta approach. Whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

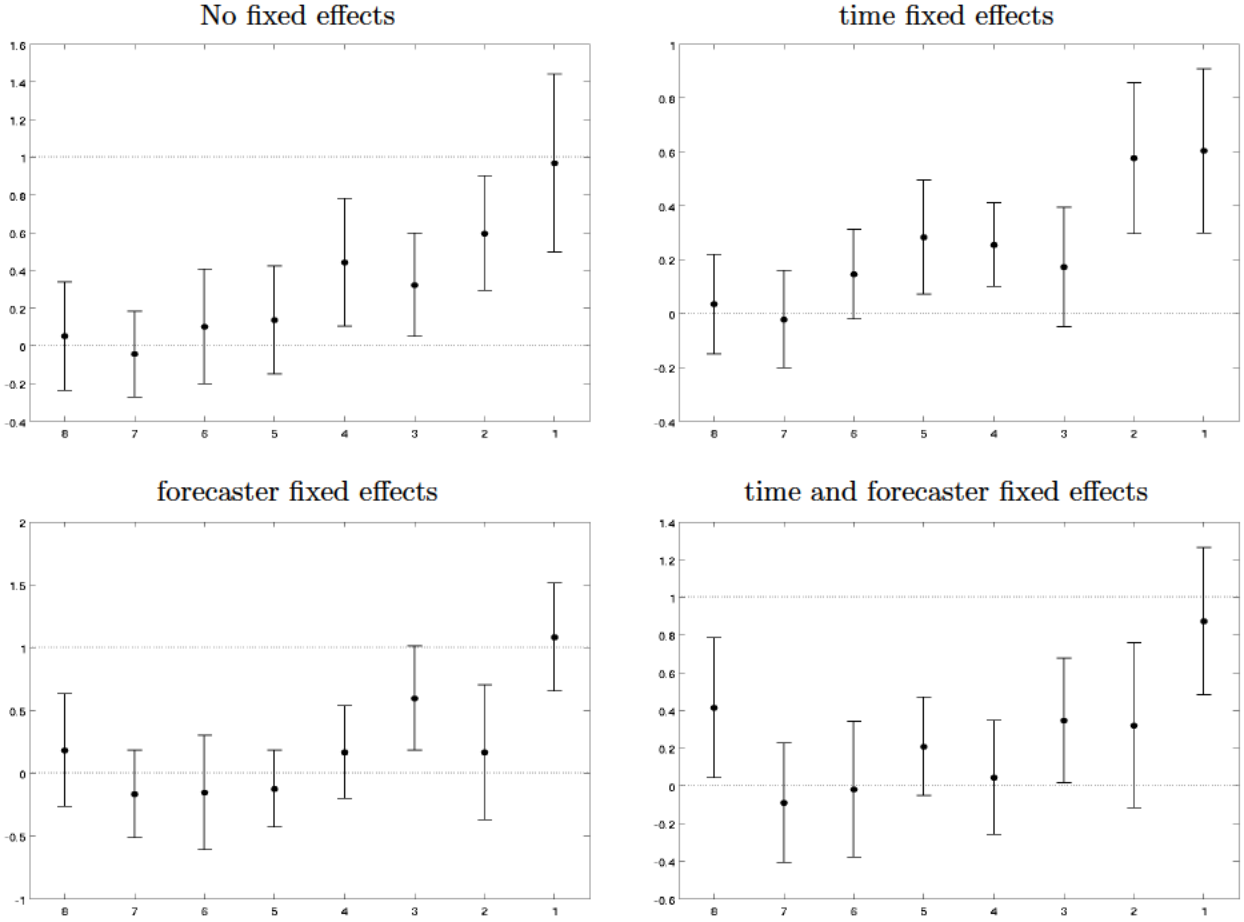
Figure A-23: A variation test: first finals vs latest vintage

Output growth Inflation



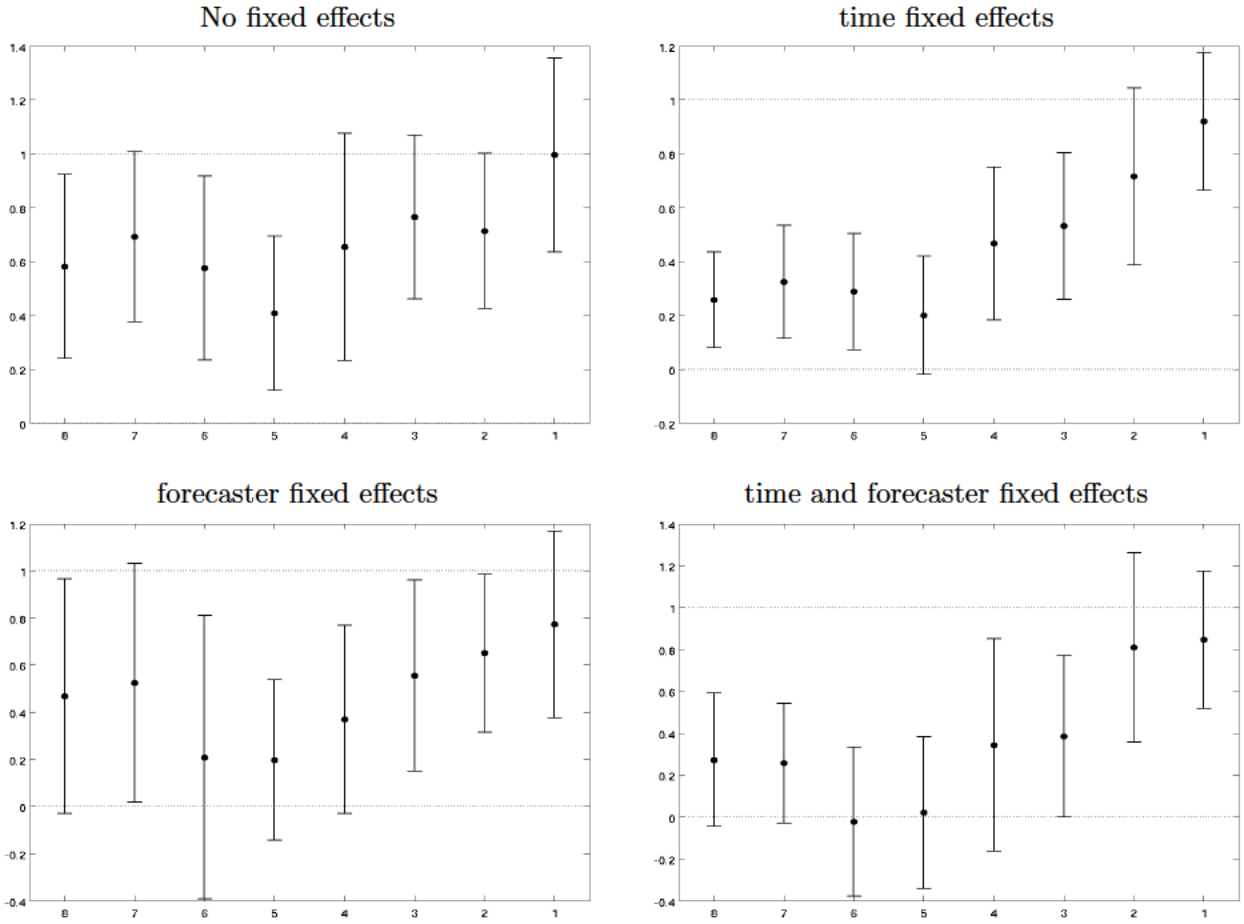
Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$ using first finals as a measure of actual realizations y_t. Gray crosses correspond to OLS estimates when the latest vintage is used instead. Whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-24: A variation test: output growth; 1982-2018 sample



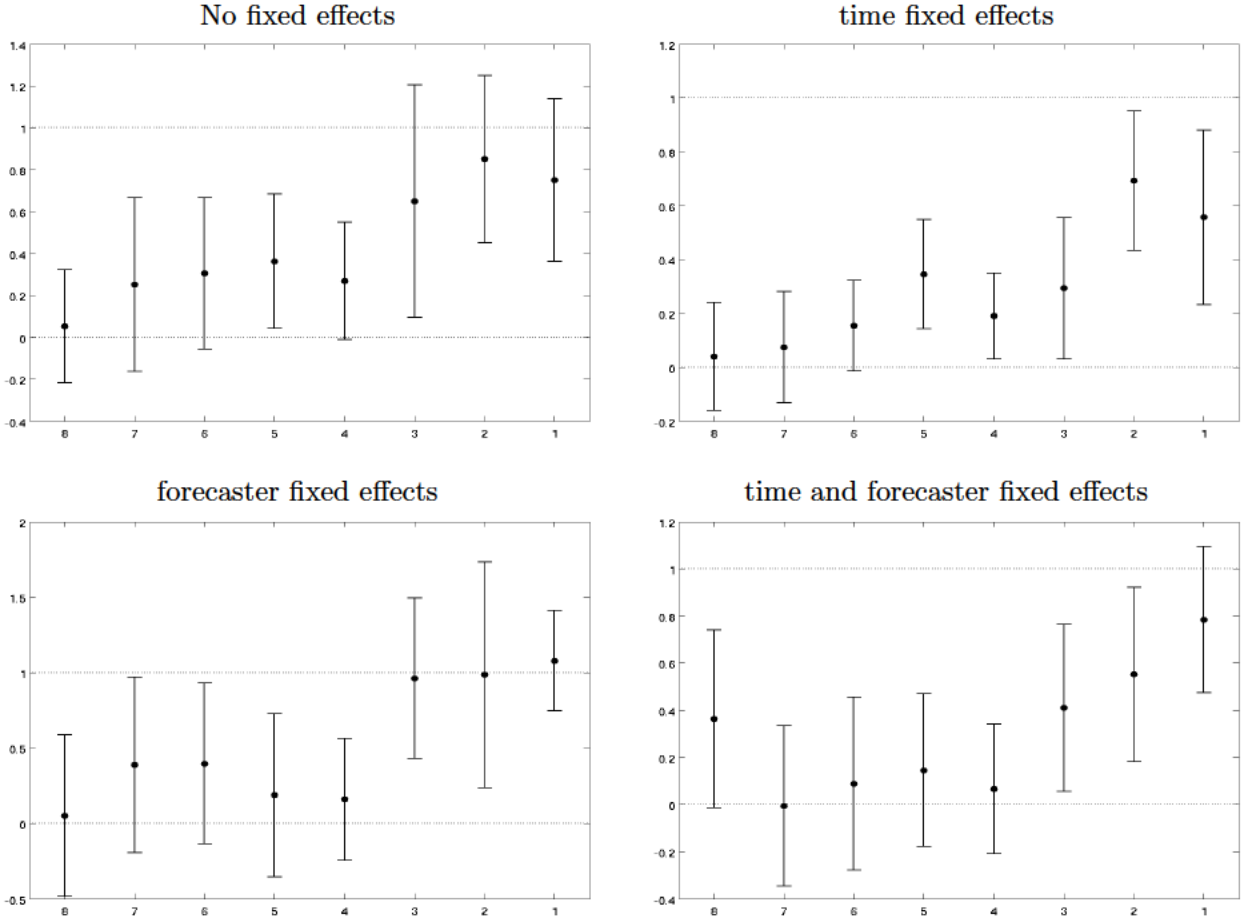
Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-25: A variation test: inflation; 1982-2018 sample



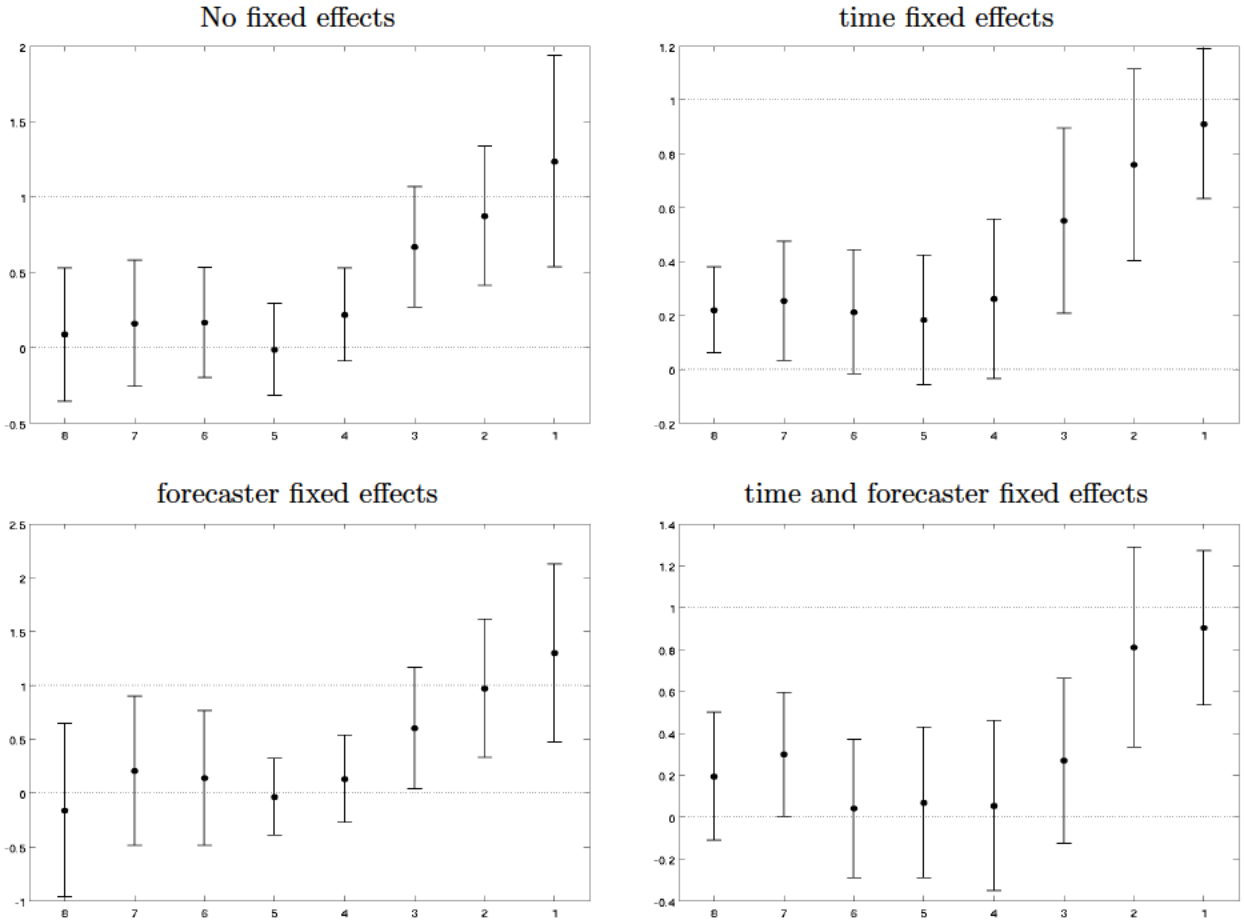
Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-26: A variation test: output growth; 1992-2022 sample



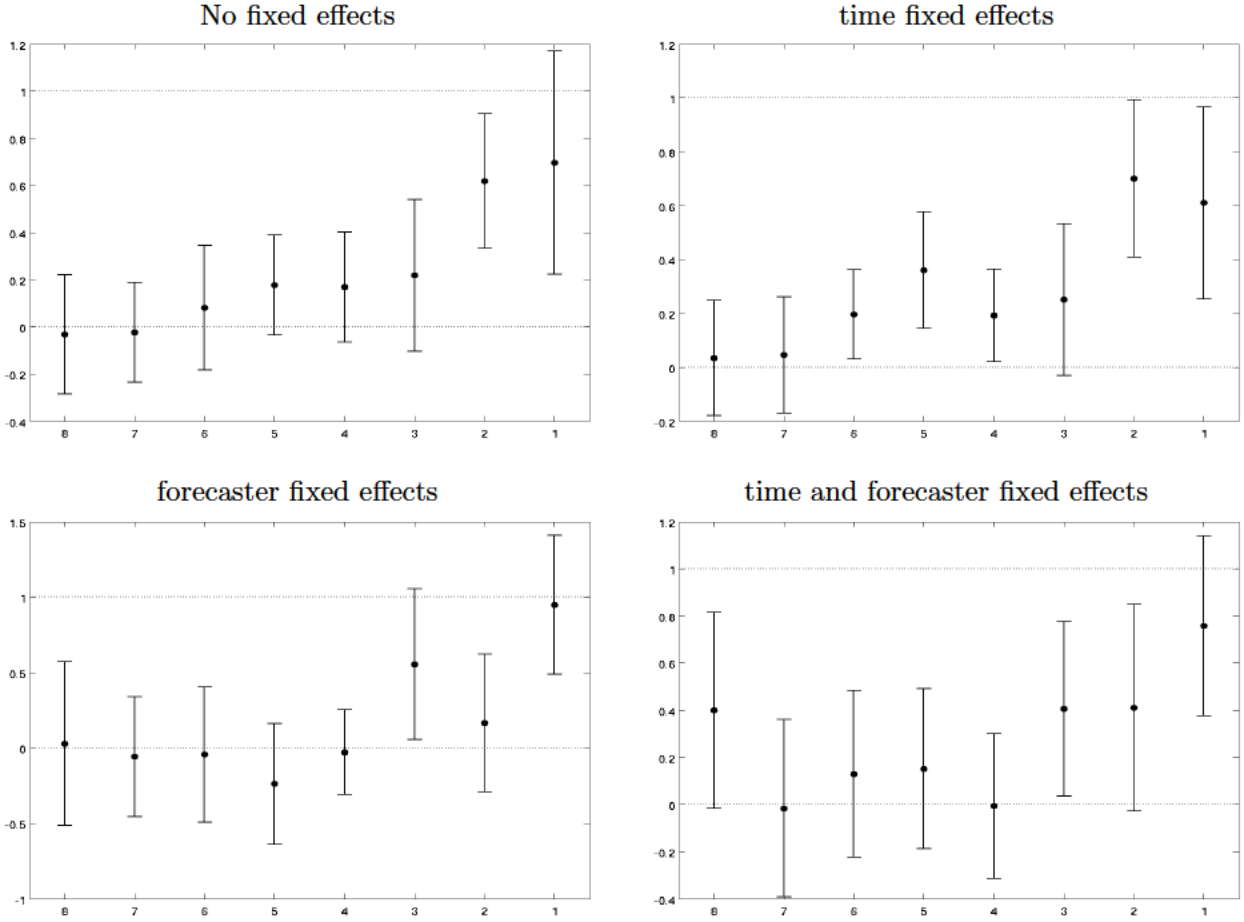
Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-27: A variation test: inflation; 1992-2022 sample



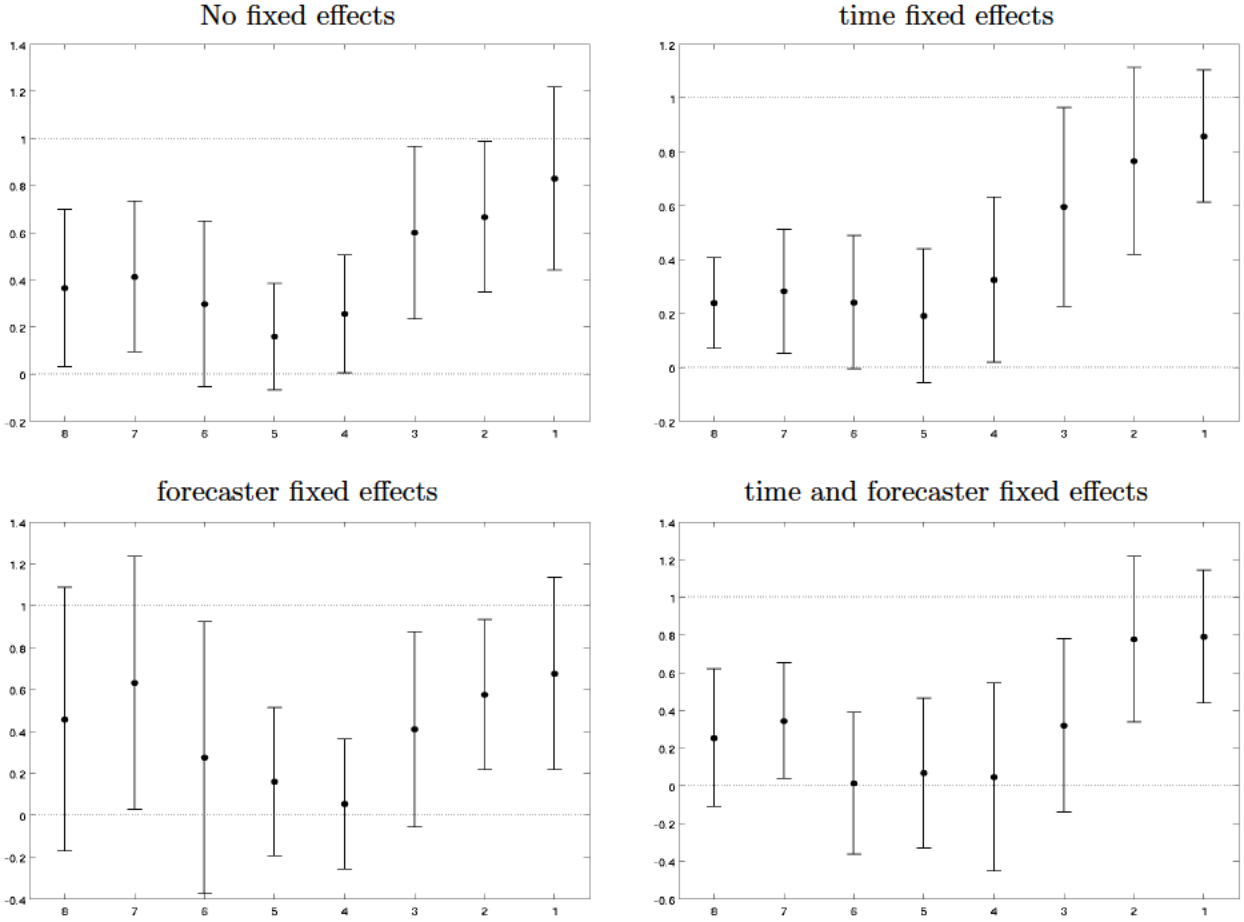
Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-28: A variation test: output growth; 1992-2018 sample



Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

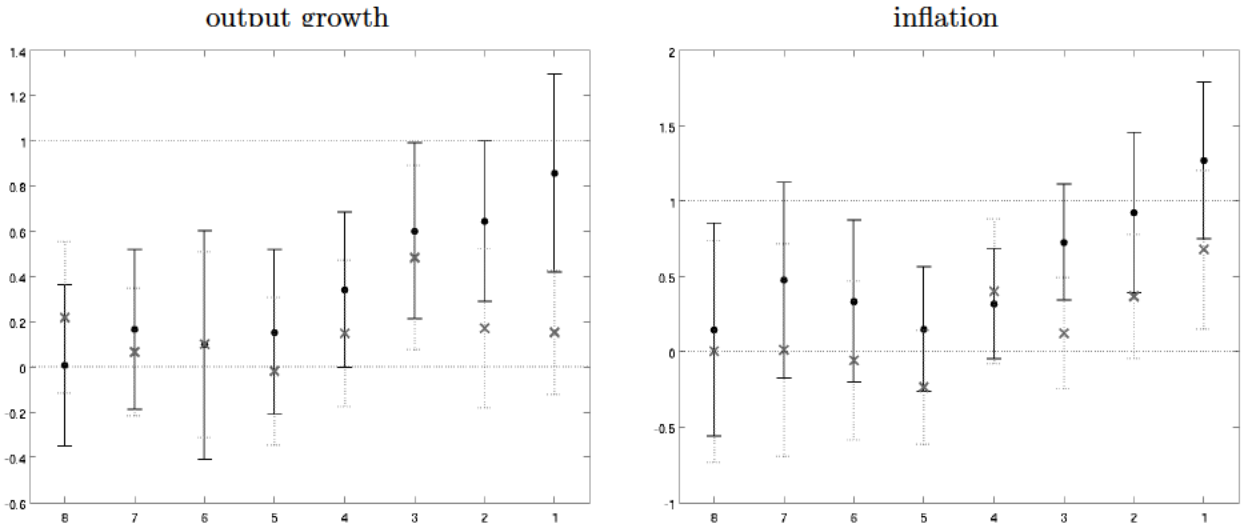
Figure A-29: A variation test: inflation; 1992-2018 sample



Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

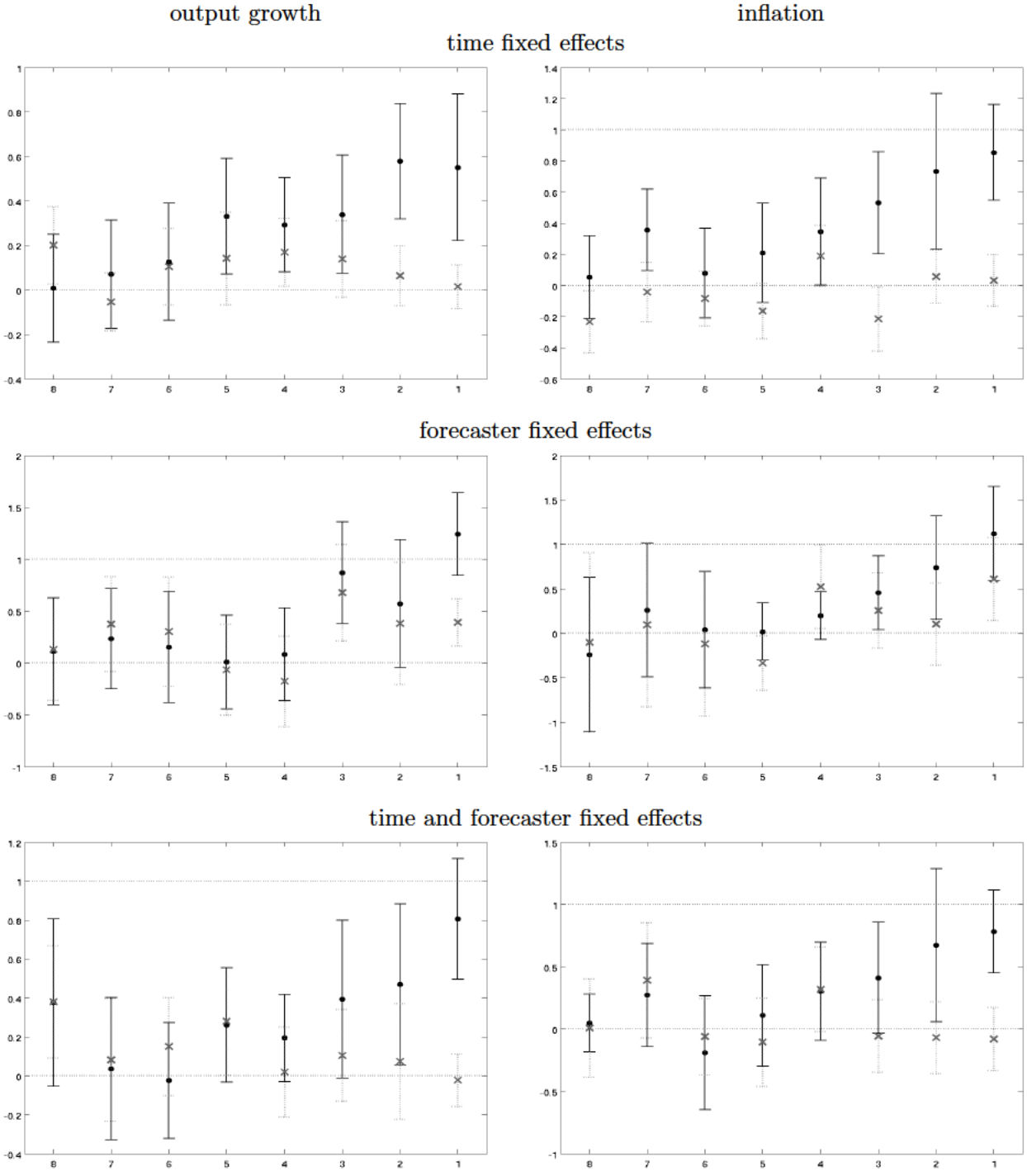
A variation test: Additional Results (weighted)

Figure A-30: A variation test: Mean vs Point Projections–weighted



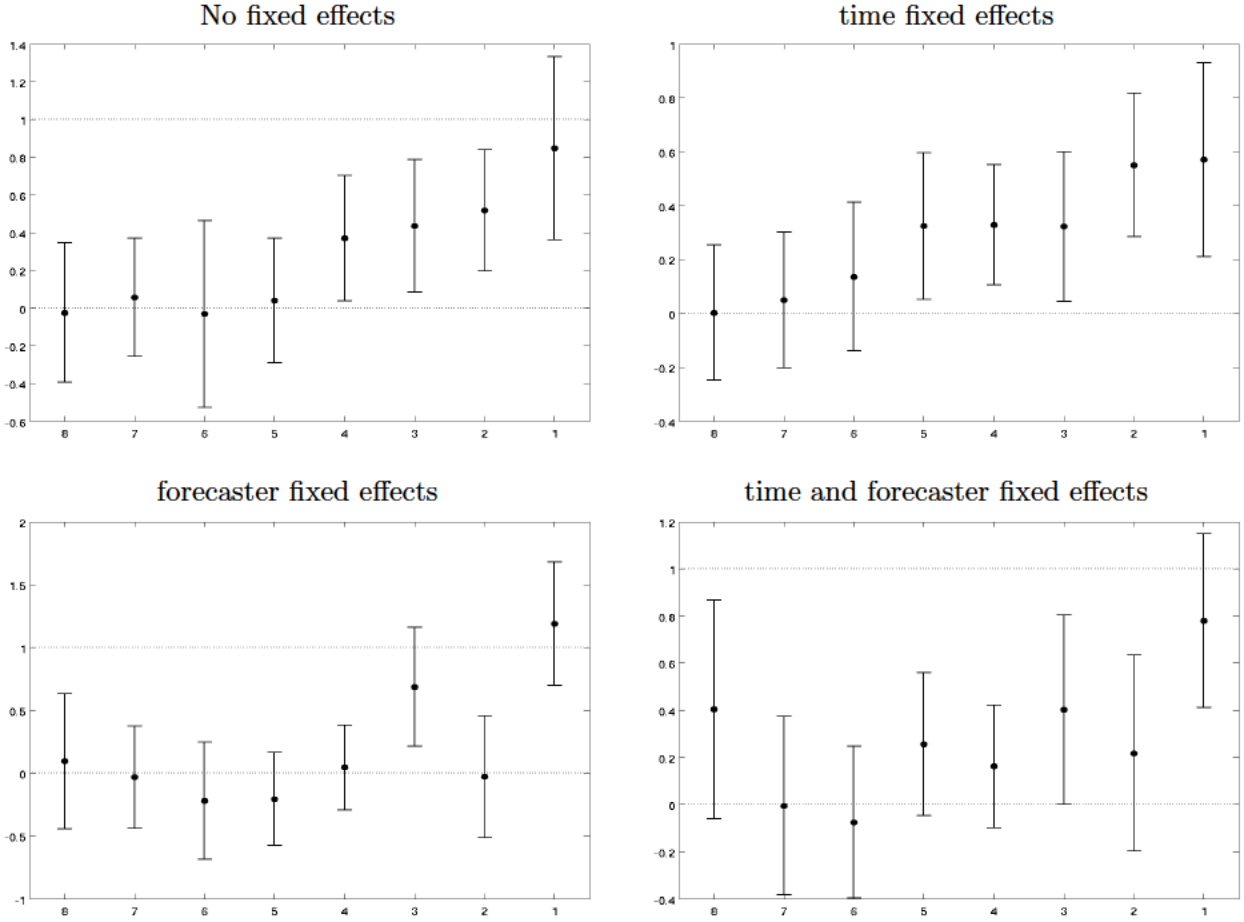
Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$. Gray crosses correspond to OLS estimates when the point prediction $y^{pp}_{t-q,i}$ is used in place of $\mathbb{E}_{t-q,i}[y_t]$. Whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-31: A variation test: regressions with fixed effects for both mean and point forecasts–weighted



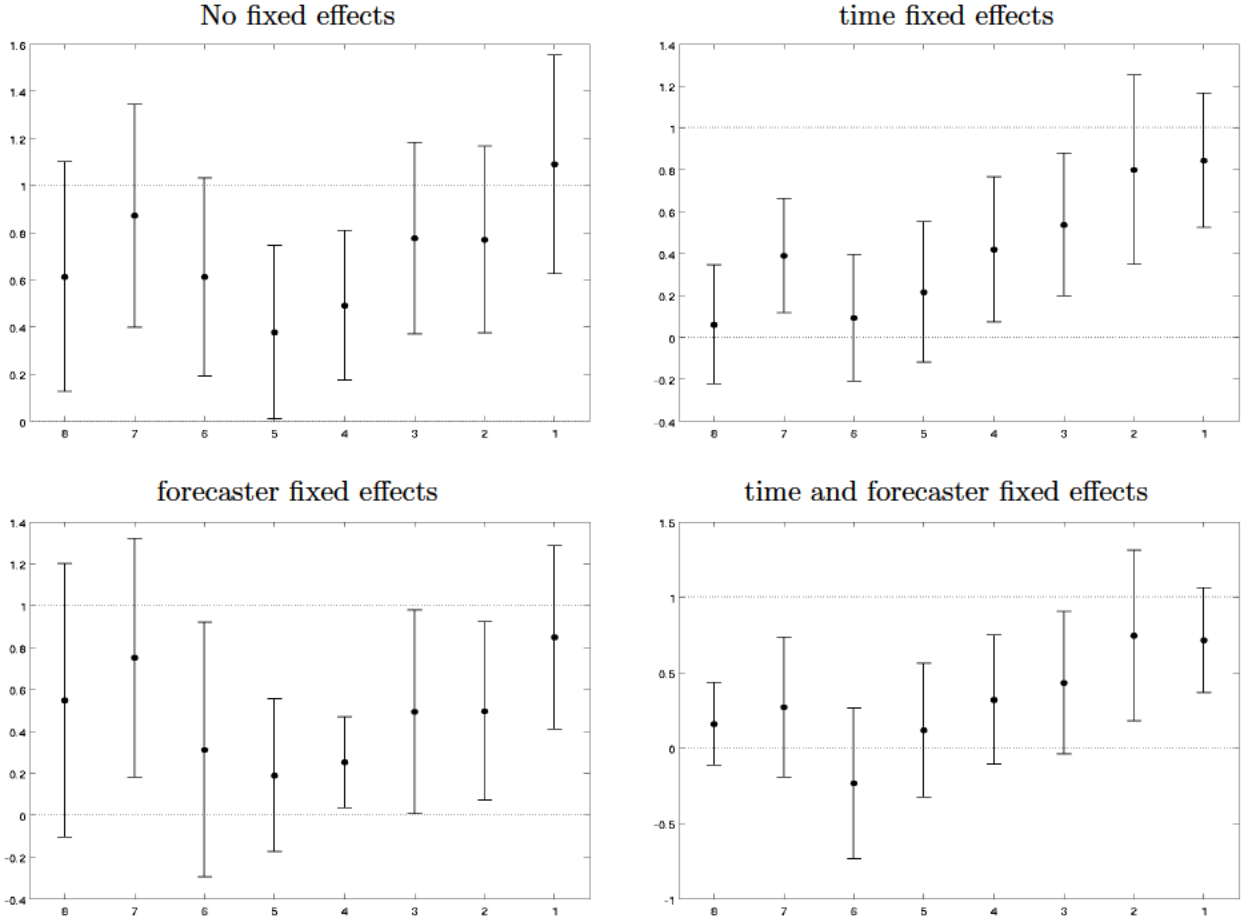
Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) using time (top panels), forecaster fixed effects (middle panels), or both (bottom panels), for $q = 8, ..., 1$. Gray crosses correspond to OLS estimates when the point prediction $y^{pp}_{t-q,i}$ is used in place of $\mathbb{E}_{t-q,i}[y_t]$. Whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-32: A variation test: output growth; 1982-2018 sample–weighted



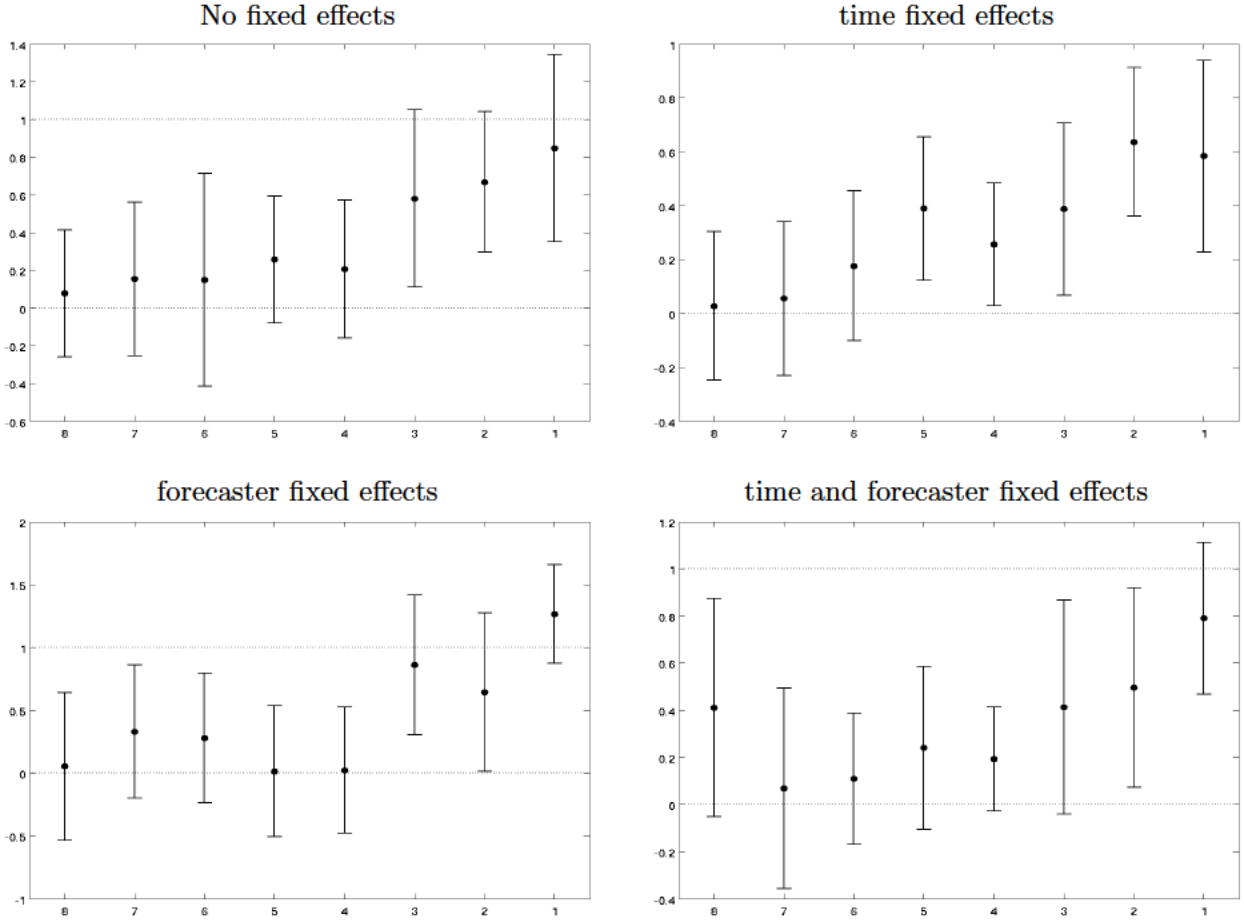
Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-33: A variation test: inflation; 1982-2018 sample–weighted



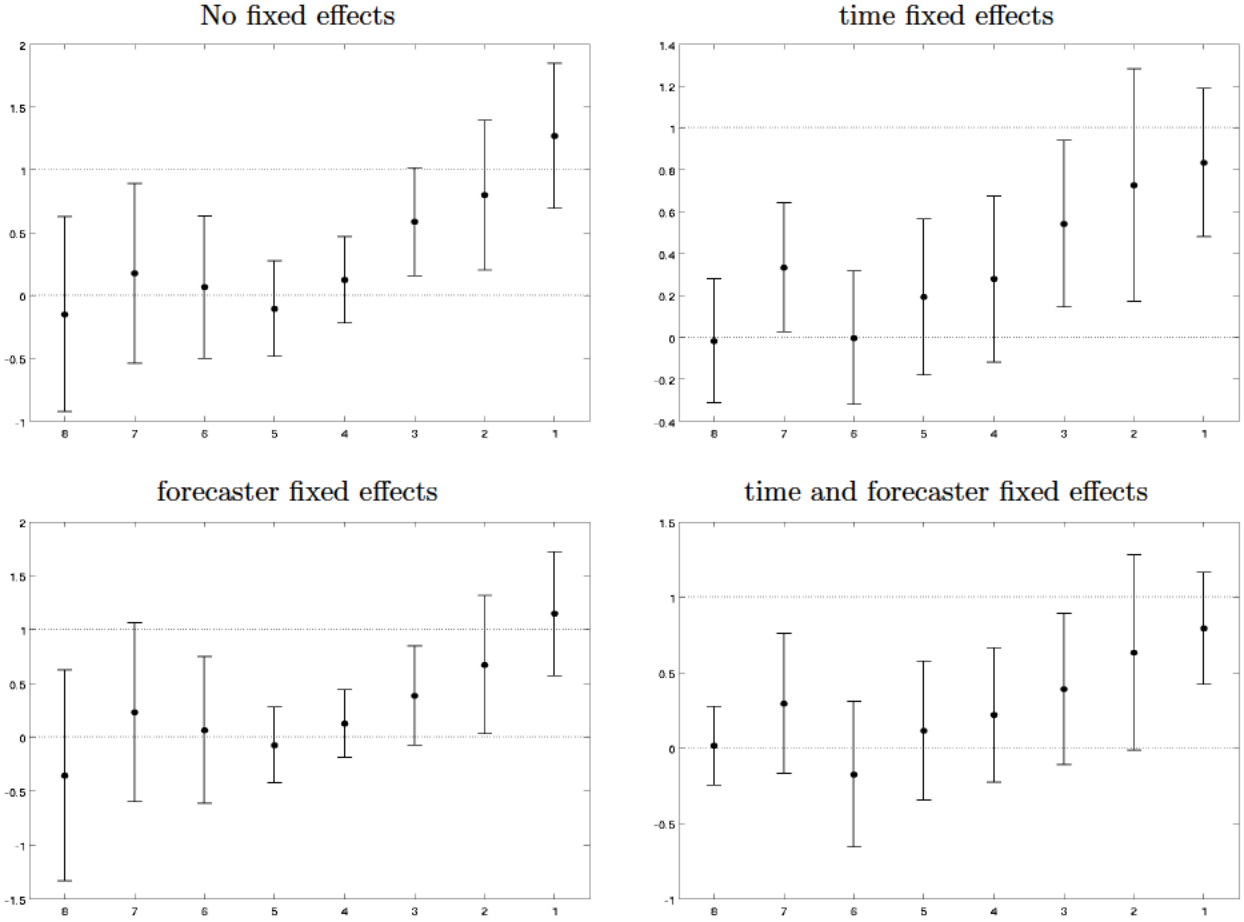
Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-34: A variation test: output growth; 1992-2022 sample–weighted



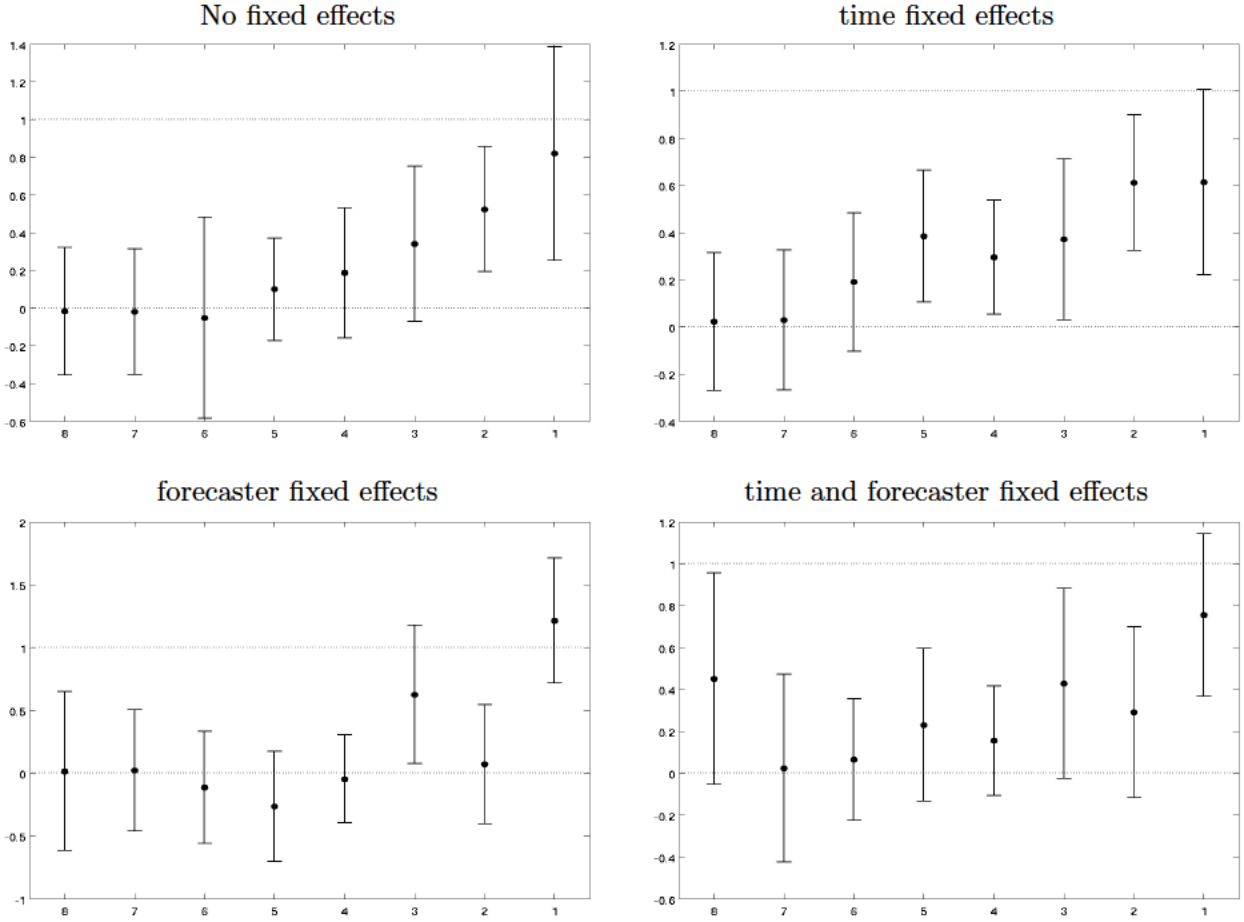
Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-35: A variation test: inflation; 1992-2022 sample–weighted



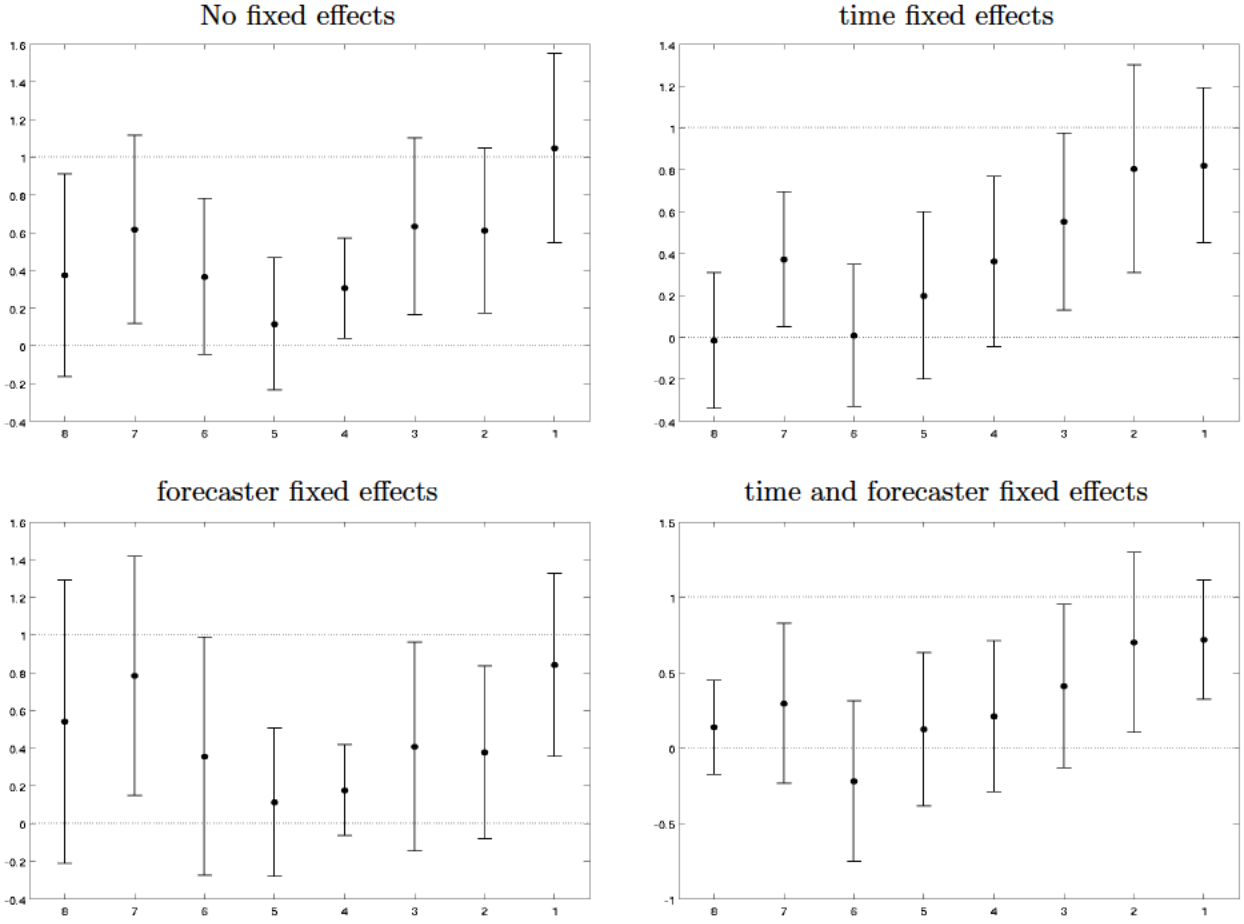
Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-36: A variation test: output growth; 1992-2018 sample–weighted



Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-37: A variation test: inflation; 1992-2018 sample–weighted



Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

A location test: additional results

Table A-6: Statistics underlying Figure 10

quarters ahead:	8	7	6	5	4	3	2	1
Period:	1982- 2022							
	Output growth							
	Log Ratio of Mean Forecast Errors							
Coef.:	-0.031	-0.057	-0.061	0.009	0.013	0.219	0.096	0.624
SE:	0.041	0.048	0.042	0.033	0.053	0.056	0.073	0.127
nobs:	1169	1271	1237	1294	1287	1371	1280	1329
no fcters:	588	588	588	589	588	591	595	592
no years:	40	40	40	40	41	41	41	41
	Fair Shiller Regressions							
Constant.:	1.823	1.167	0.579	-0.370	0.004	0.324	0.284	0.013
SE:	0.731	0.830	0.886	0.643	0.346	0.233	0.162	0.033
Coef mean:	0.141	0.257	0.404	0.209	0.211	0.061	0.208	0.025
SE:	0.156	0.111	0.170	0.102	0.041	0.058	0.088	0.012
Coef. pp:	0.110	0.254	0.331	0.889	0.764	0.801	0.699	0.973
SE:	0.072	0.187	0.182	0.178	0.075	0.083	0.062	0.018
nobs:	1169	1271	1237	1294	1287	1371	1280	1329
no fcters:	588	588	588	589	588	591	595	592
no years:	40	40	40	40	41	41	41	41
R2:	0.020	0.060	0.120	0.400	0.530	0.880	0.930	0.980
	Inflation							
	Log Ratio of Mean Forecast Errors							
Coef.:	0.075	0.160	0.087	0.132	0.193	0.416	0.569	1.219
SE:	0.048	0.067	0.057	0.055	0.084	0.070	0.098	0.126
nobs:	1133	1235	1196	1254	1240	1316	1243	1288
no fcters:	588	588	588	589	588	591	595	592
no years:	40	40	40	40	41	41	41	41
	Fair Shiller Regressions							
Constant.:	1.411	1.275	1.016	0.738	0.855	0.198	0.167	0.091
SE:	0.366	0.365	0.276	0.211	0.318	0.104	0.073	0.073
Coef mean:	0.227	0.254	0.336	0.275	0.237	0.143	0.014	0.088
SE:	0.065	0.104	0.063	0.088	0.130	0.060	0.036	0.061
Coef. pp:	0.127	0.142	0.157	0.364	0.442	0.753	0.904	0.868
SE:	0.054	0.073	0.061	0.095	0.133	0.087	0.061	0.084
nobs:	1133	1235	1196	1254	1240	1316	1243	1289
no fcters:	588	588	588	589	588	591	595	592
no years:	40	40	40	40	41	41	41	41
R2:	0.150	0.190	0.260	0.350	0.510	0.830	0.940	0.950

E Robustness to Priors and the $F(\cdot)$ Function

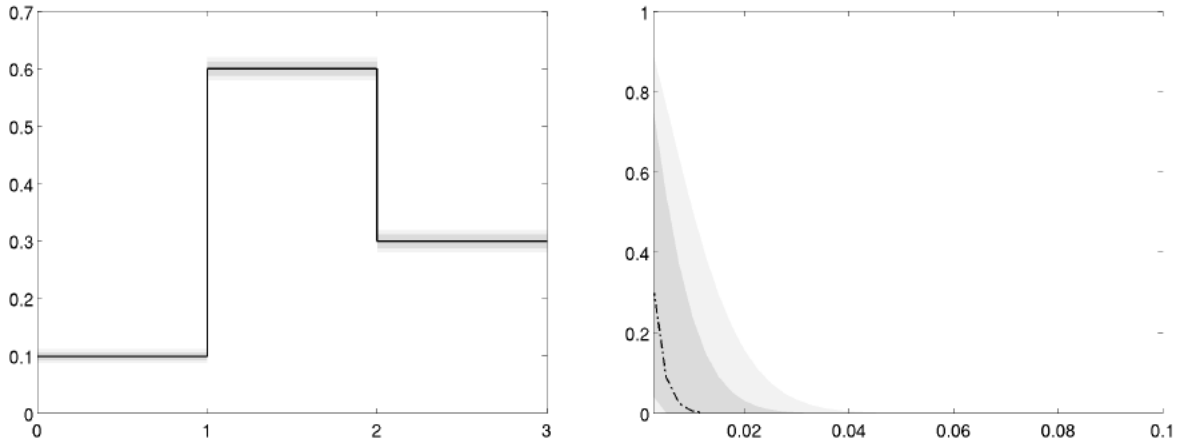
E.A "Halving the noise" prior

Under this prior we assume that the parameters a_ϕ and b_ϕ of the truncated gamma distribution $\mathcal{G}a(a_\phi, b_\phi)\mathbb{I}_{(\underline{\phi},\infty)}(\phi)$ defining the prior for the precision parameter ϕ determining the amount of noise are such that at the prior mean, when the underlying probability ν is 50 percent, the standard deviation of the noise is 1.25 percent, as opposed to 2.5 percent in the baseline estimation. The parameters a_ϵ and b_ϕ of the truncated gamma distribution $\mathcal{G}a(a_\epsilon, b_\epsilon)\mathbb{I}_{(0,\bar{\epsilon})}(\epsilon)$ under this prior are such that at the prior mean the probability of reporting a zero, ϱ, is 2.5 percent, as opposed to 5 percent in the baseline, when the underlying ν is 2.5 percent.

The comparison of Figure A-38 below and Figure A-2 provides a visual idea of the differences in the priors. As a reminder, the left panel of Figure A-38 shows the 50 and 90 percent a-priori coverage intervals for the noise associated with three different values of ν: 0.1, 0.6 and 0.3, while the right panel shows the mean, the 50, and the 90 percent coverage intervals of ϱ as a function of ν.

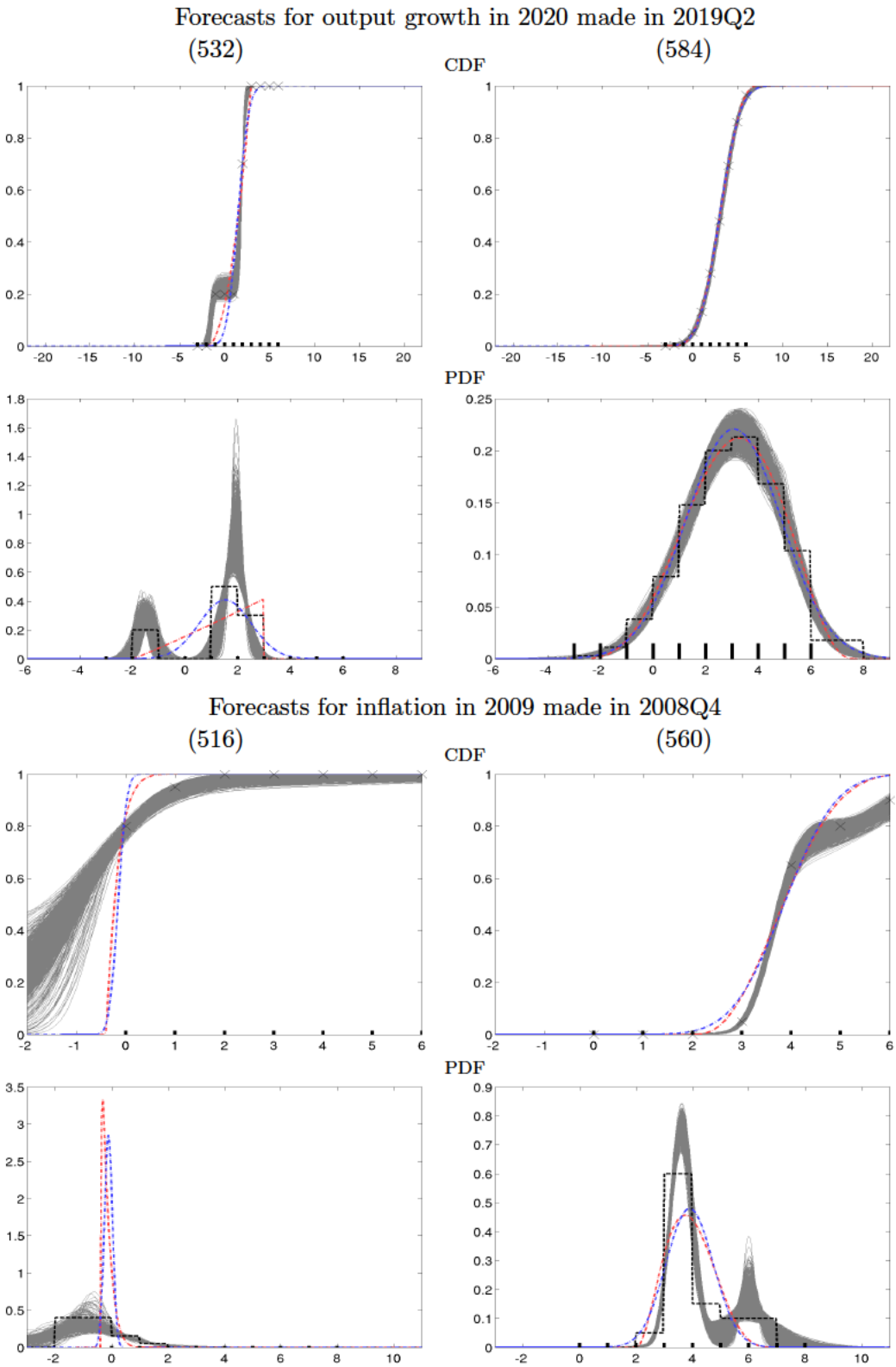
Figure A-38: Noise and zero probability under "halving the noise" prior

Noise around ν $\qquad$ $\varrho(\nu,\epsilon) = \int_0^\epsilon b(x|\nu,r)dx$ as a function of ν



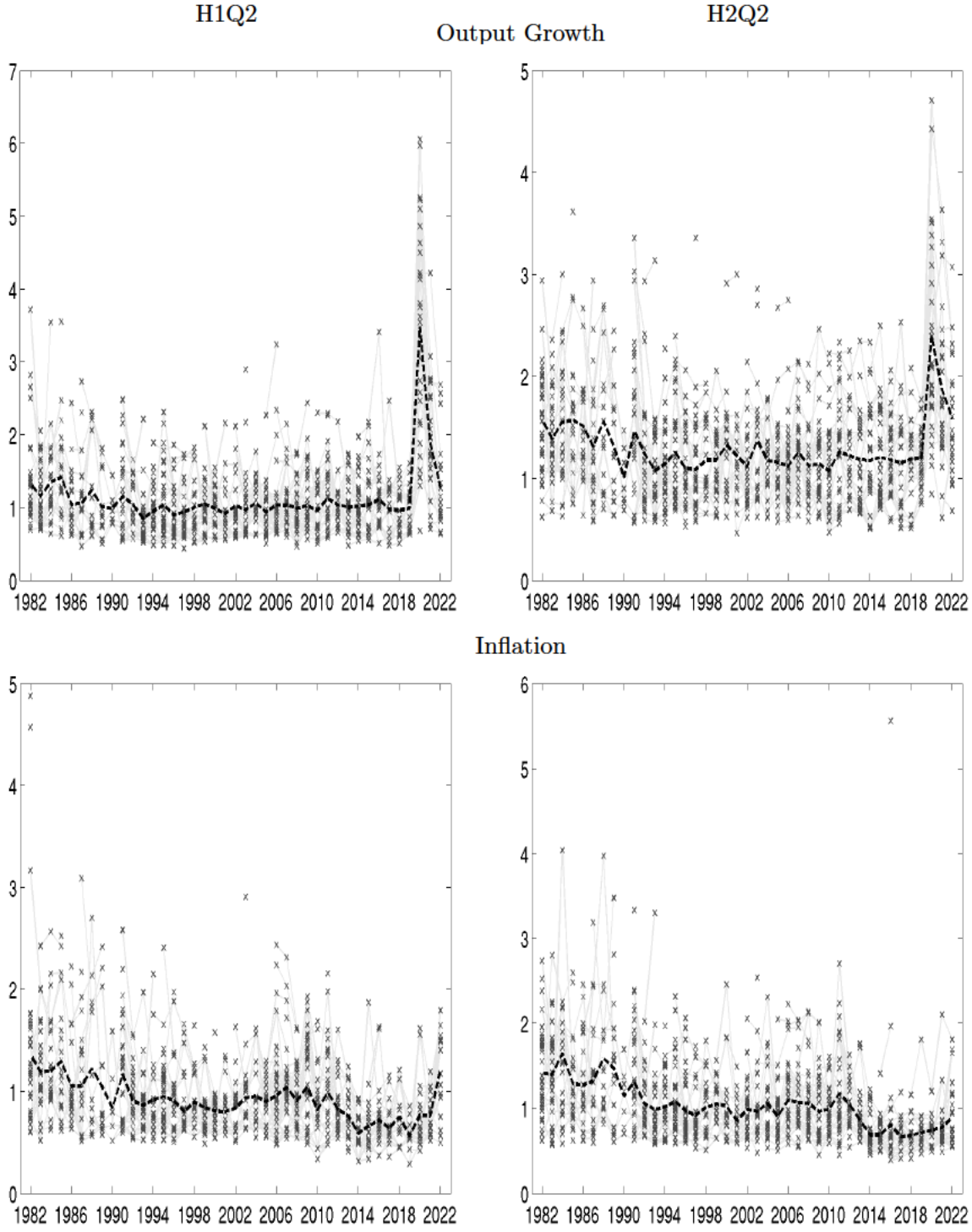
Note: The left panel shows the 50 and 90 percent a-priori coverage intervals for the noise associated with three different values of ν: 0.1, 0.6 and 0.3.

Figure A-39: Inference using the Bayesian nonparametric approach: CDFs and PDFs for selected examples [Figure 2 under "halving the noise" prior]



Note: For each forecaster the top and bottom panels show the subjective CDF and PDF, respectively, estimated using the BNP approach (posterior draws; light gray), as well as the least-squares estimates obtained under the normal (gray, dashed line) and the beta (black, dash-and-dotted line) parametric assumptions. The CDF panels also show the observed cumulated histogram probabilities Z_{ij} $j = 1, \ldots, J$ (crosses), while the PDF panel show the step-wise uniform PDF (gray dashed lines) implied by such probabilities.

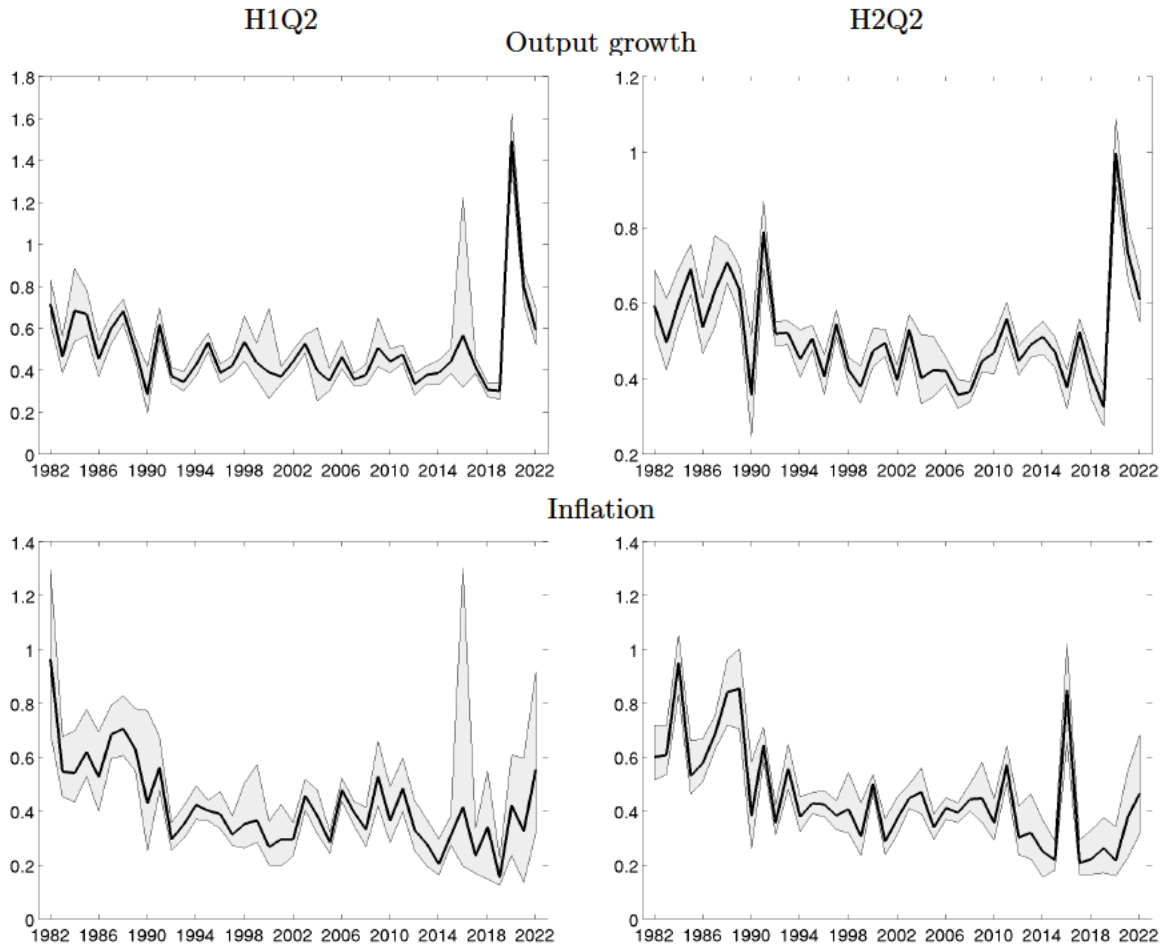
Figure A-40: Subjective uncertainty by individual respondent [Figure 3 under "halving the noise" prior]



Note: Each panel displays the posterior mean of the standard deviation of the subjective predictive distribution by individual respondent (light gray crosses, connected by thin gray line whenever the respondent appears in consecutive surveys), and the cross-sectional average of the individual standard deviations (dashed black line). Top and bottom panels correspond to output growth and inflation projections. The left and right column correspond to current and next year projections.

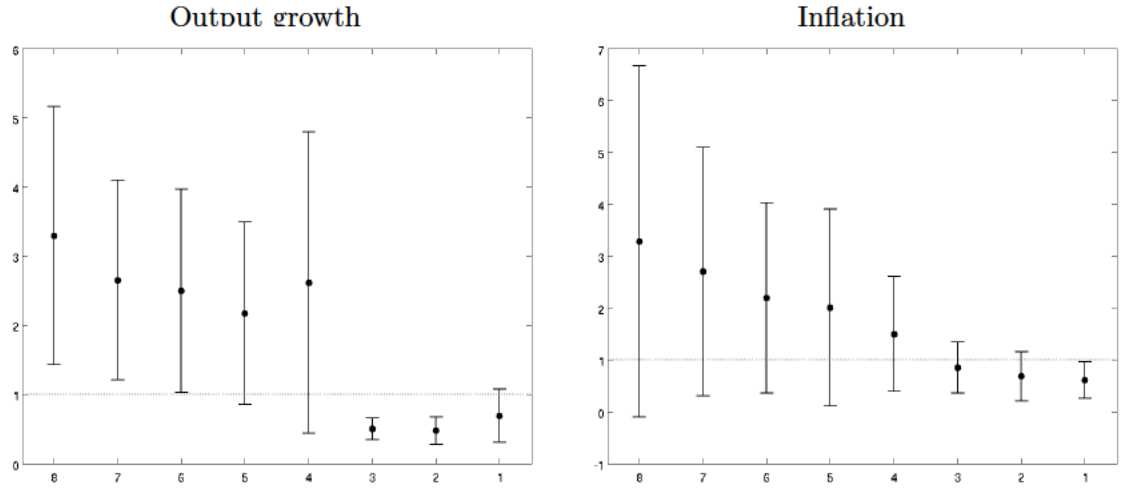
Figure A-41: Cross-sectional standard deviation of individual uncertainty—Q2 survey

[Figure 4 under "halving the noise" prior]



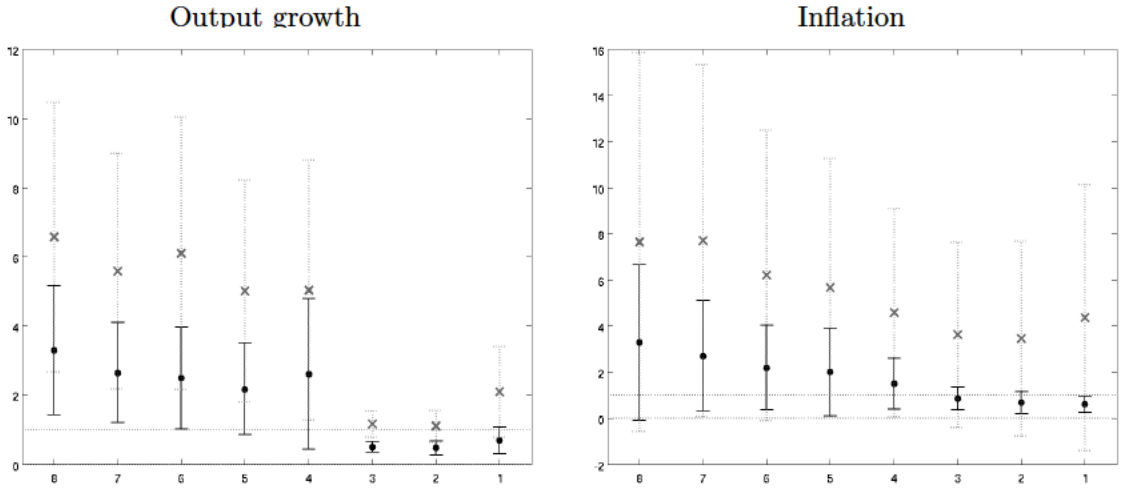
Note: Posterior mean of the cross-sectional standard deviation of the individual standard deviations (solid black line) The shaded areas display the 90 percent posterior credible intervals. Top panels: output growth. Bottom panels: inflation. Left column: current year; right column: next year.

Figure A-42: Do forecasters over or under-estimate uncertainty? A scale test [Figure 5 under "halving the noise" prior]



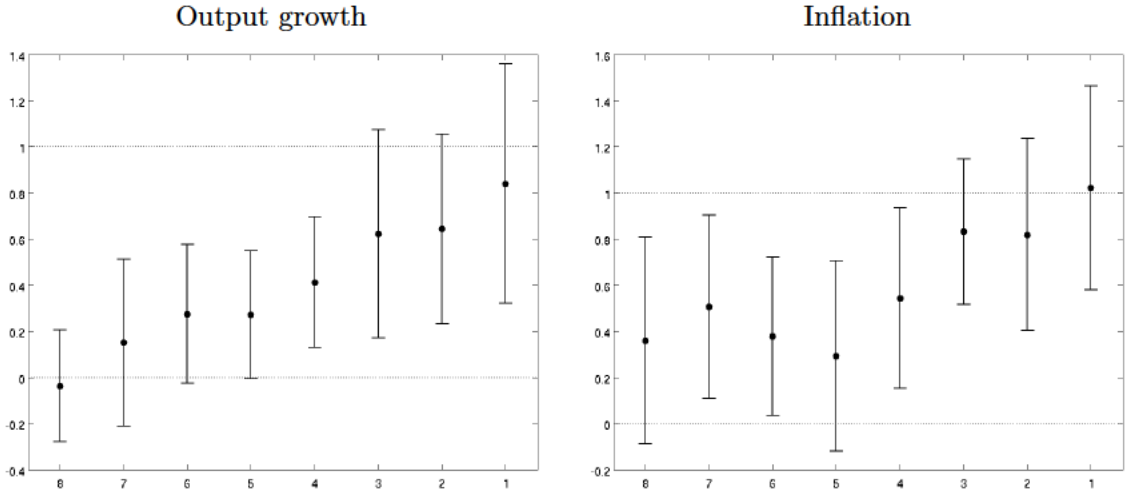
Note: Black dots correspond to OLS estimates of α_q from regression (18) for $q = 8, .., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-43: A scale test—comparison with the generalized beta approach [Figure 6 under "halving the noise" prior]



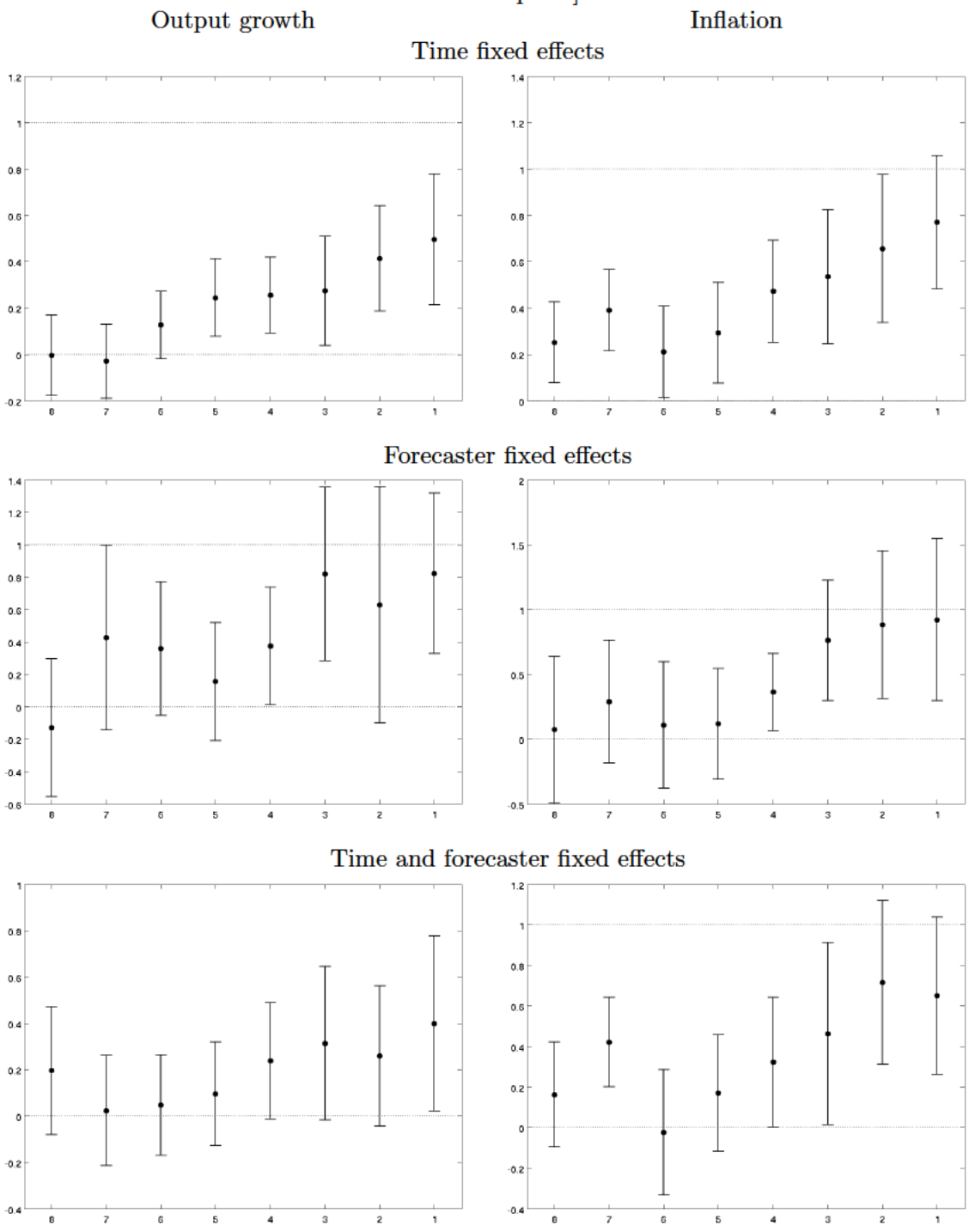
Note: Black dots correspond to OLS estimates of α_q from regression (18) for $q = 8, ..., 1$ using the posterior means for $\mathbb{E}_{t-q,i}[y_t]$ and $\sigma^2_{t|t-q,i}$ from our approach. Gray crosses correspond to OLS estimates when these objects are obtained using the generalized beta approach. Whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-44: Do differences in subjective uncertainty map into differences in forecast accuracy? A variation test [Figure 7 under "halving the noise" prior]



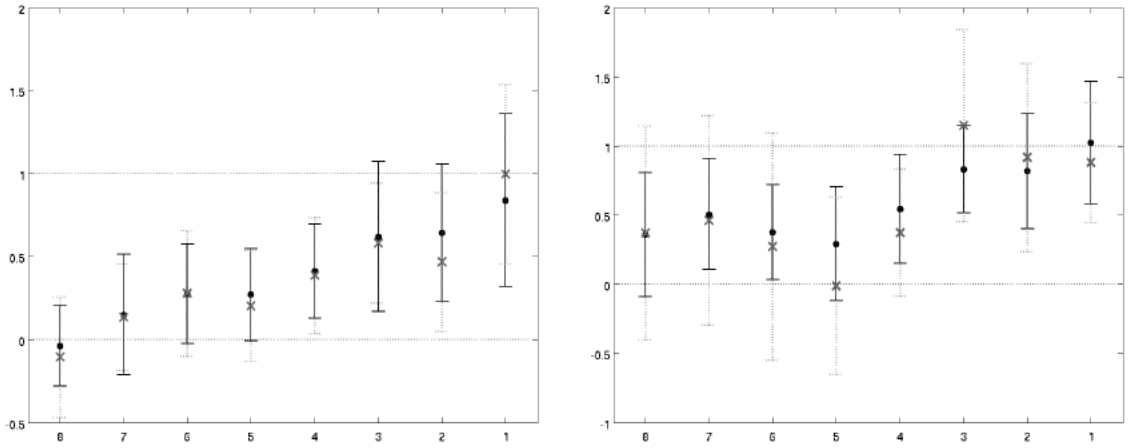
Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, .., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-45: A variation test—regressions with fixed effects [Figure 8 under "halving the noise" prior]



Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) using time (top panels), forecaster fixed effects (middle panels), or both (bottom panels), for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-46: A variation test—accounting for inference uncertainty (baseline vs weighted OLS) [Figure 9 under "halving the noise" prior]

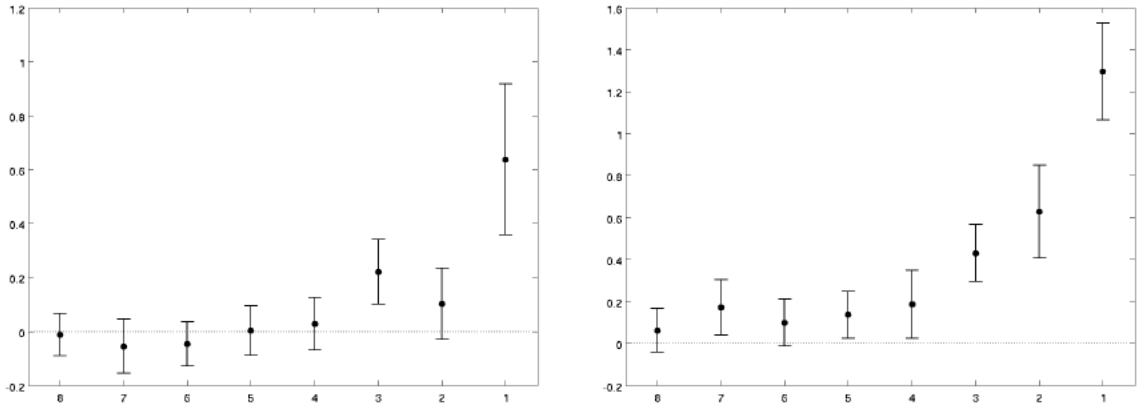


Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors. Gray crosses correspond to weighted OLS estimates, where the weights are inversely proportional to inference uncertainty as measured by the interquantile range of the posterior distribution of $\sigma_{t|t-q,i}$. Whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

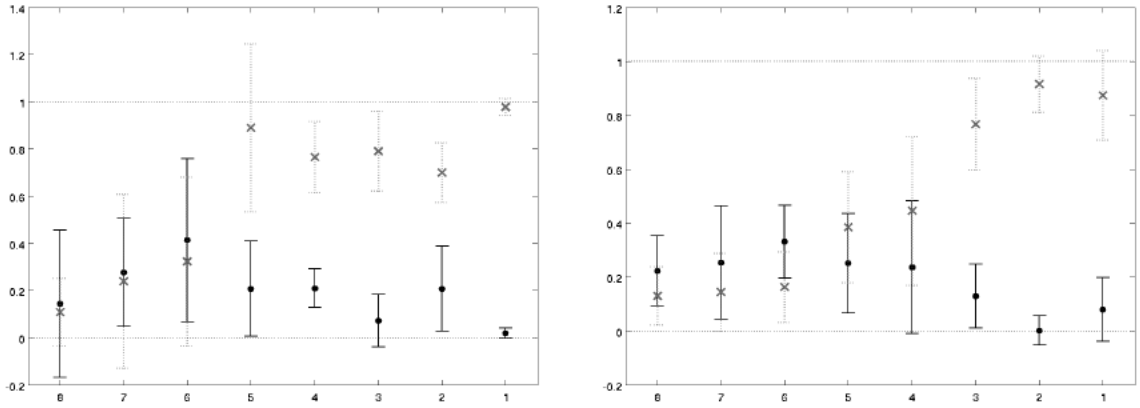
Figure A-47: A location test: Relative accuracy of mean vs point projections [Figure 10 under "halving the noise" prior]

Output growth Inflation

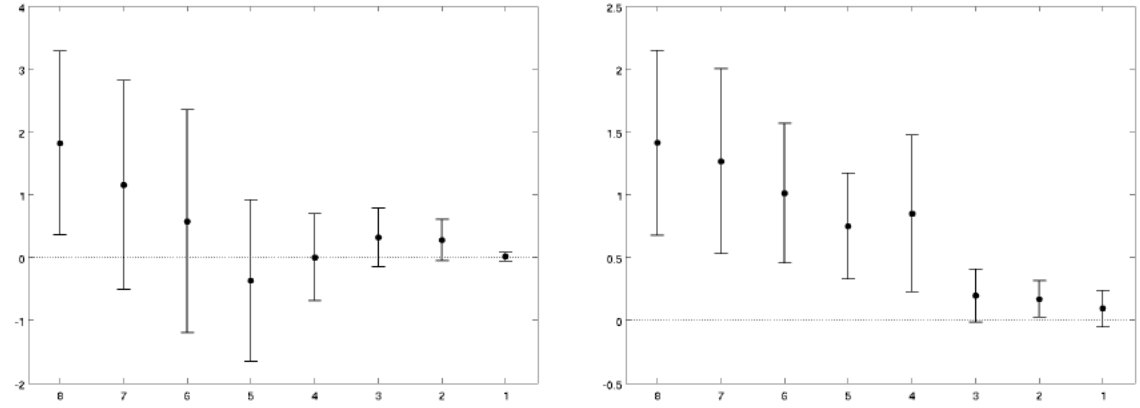
Log Ratio of Squared Forecast Errors (γ_q)



Fair-Shiller Regressions: Coefficients on Mean ($\delta_{1,q}$, Black) and Point Prediction ($\delta_{2,q}$, Gray)



Fair-Shiller Regressions: Intercept $\delta_{0,q}$



Note: Top panel: Black dots correspond to OLS estimates of γ_q from regression (20) for $q = 8, ..., 1$. Middle panel: Black dots and gray crosses correspond to OLS estimates of $\delta_{1,q}$ and $\delta_{2,q}$, respectively, from regression (21) for $q = 8, ..., 1$. Bottom panel: Black dots correspond to OLS estimates of the constant $\delta_{0,q}$ from regression (21) for $q = 8, ..., 1$. In all panels, whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

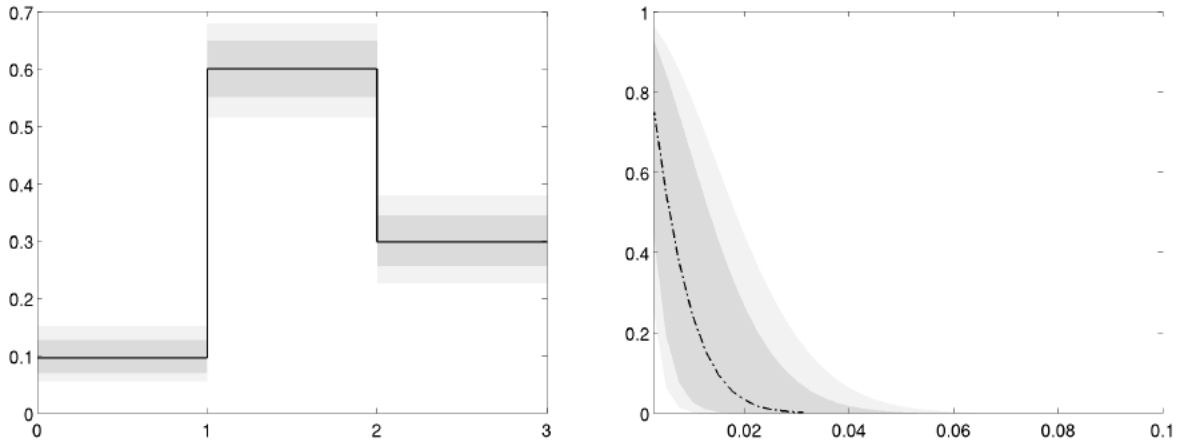
E.B "Doubling the noise" prior

Under this prior we assume that the parameters a_ϕ and b_ϕ of the truncated gamma distribution $\mathcal{G}a(a_\phi, b_\phi)\mathcal{I}(\phi)_{(\underline{\phi},\infty)}$ defining the prior for the precision parameter ϕ determining the amount of noise are such that at the prior mean, when the underlying probability ν is 50 percent, the standard deviation of the noise is 5 percent, as opposed to 2.5 percent in the baseline estimation. $\underline{\phi}$ is chosen so that the noise has an upper bound of 10 percent, as opposed to 2.5 percent in the baseline estimation. The parameters a_ϵ and b_ϕ of the truncated gamma distribution $\mathcal{G}a(a_\epsilon, b_\epsilon)\mathcal{I}(\epsilon)_{(0,\bar{\epsilon})}$ under this prior are such that at the prior mean the probability of reporting a zero, ϱ, is 10 percent, as opposed to 5 percent in the baseline, when the underlying ν is 2.5 percent. $\bar{\epsilon}$ is such that ϱ is at most 40 percent for ν equal to 2.5 percent, as opposed to 20 percent in the baseline estimation.

The comparison of Figure A-48 below and Figure A-2 provides a visual idea of the differences in the priors. As a reminder, the left panel of Figure A-48 shows the 50 and 90 percent a-priori coverage intervals for the noise associated with three different values of ν: 0.1, 0.6 and 0.3, while the right panel shows the mean, the 50, and the 90 percent coverage intervals of ϱ as a function of ν.

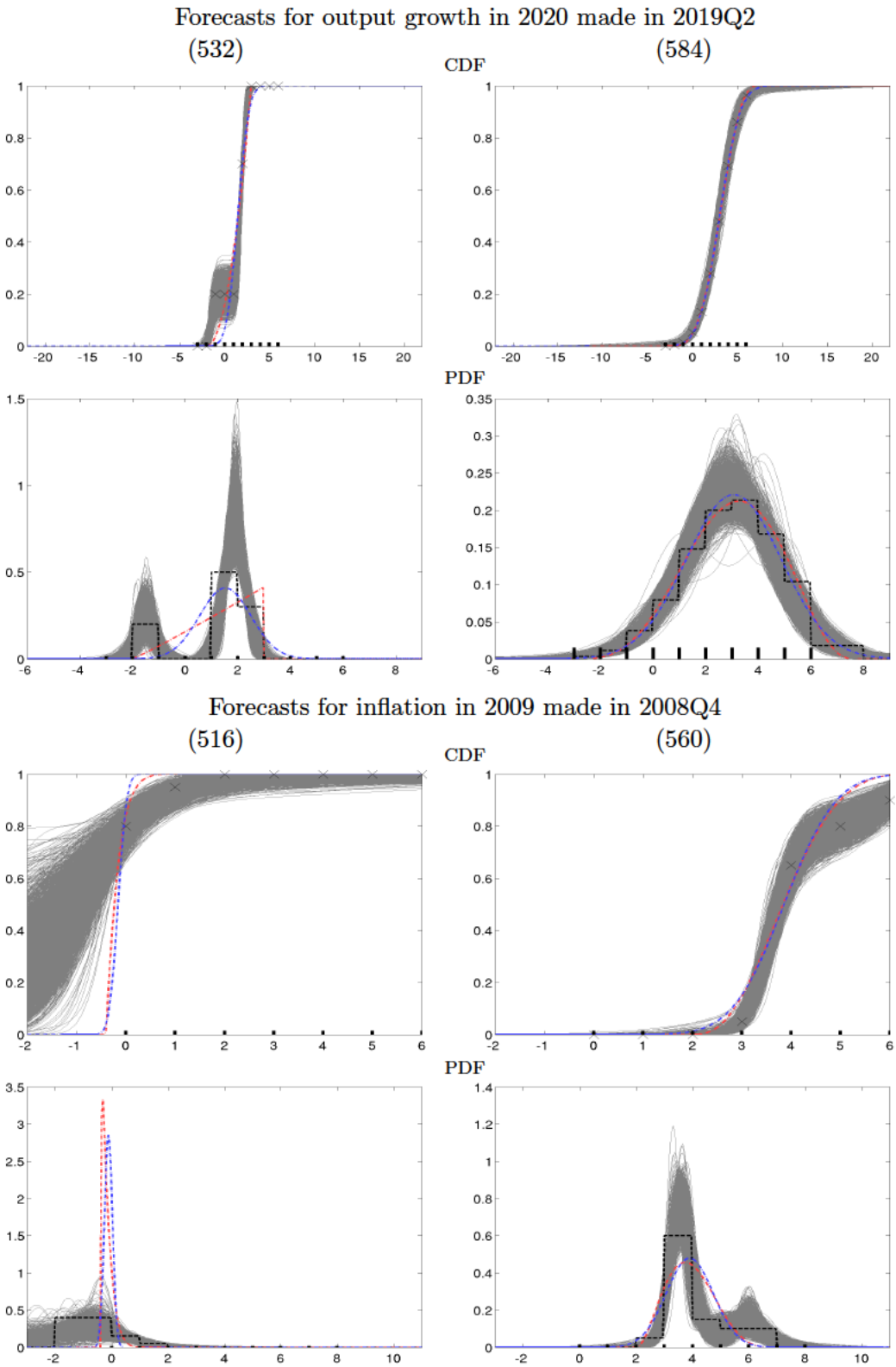
Figure A-48: Noise and zero probability under "doubling the noise" prior

Noise around ν $\qquad$ $\varrho(\nu,\epsilon) = \int_0^\epsilon b(x|\nu,r)dx$ as a function of ν



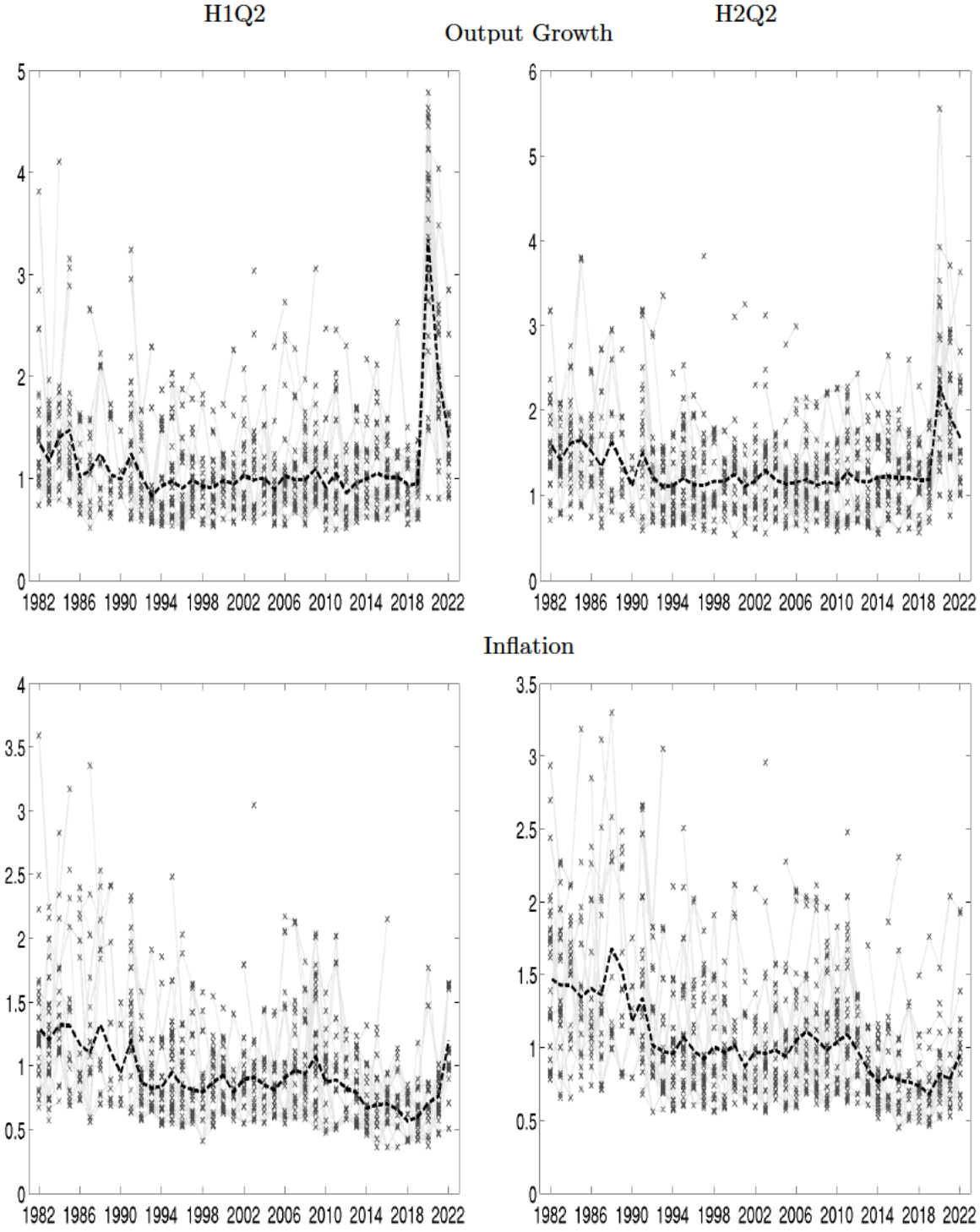
Note: The left panel shows the 50 and 90 percent a-priori coverage intervals for the noise associated with three different values of ν: 0.1, 0.6 and 0.3.

Figure A-49: Inference using the Bayesian nonparametric approach: CDFs and PDFs for selected examples [Figure 2 under "doubling the noise" prior]



Note: For each forecaster the top and bottom panels show the subjective CDF and PDF, respectively, estimated using the BNP approach (posterior draws; light gray), as well as the least-squares estimates obtained under the normal (gray, dashed line) and the beta (black, dash-and-dotted line) parametric assumptions. The CDF panels also show the observed cumulated histogram probabilities Z_{ij} $j = 1, \ldots, J$ (crosses), while the PDF panel show the step-wise uniform PDF (gray dashed lines) implied by such probabilities.

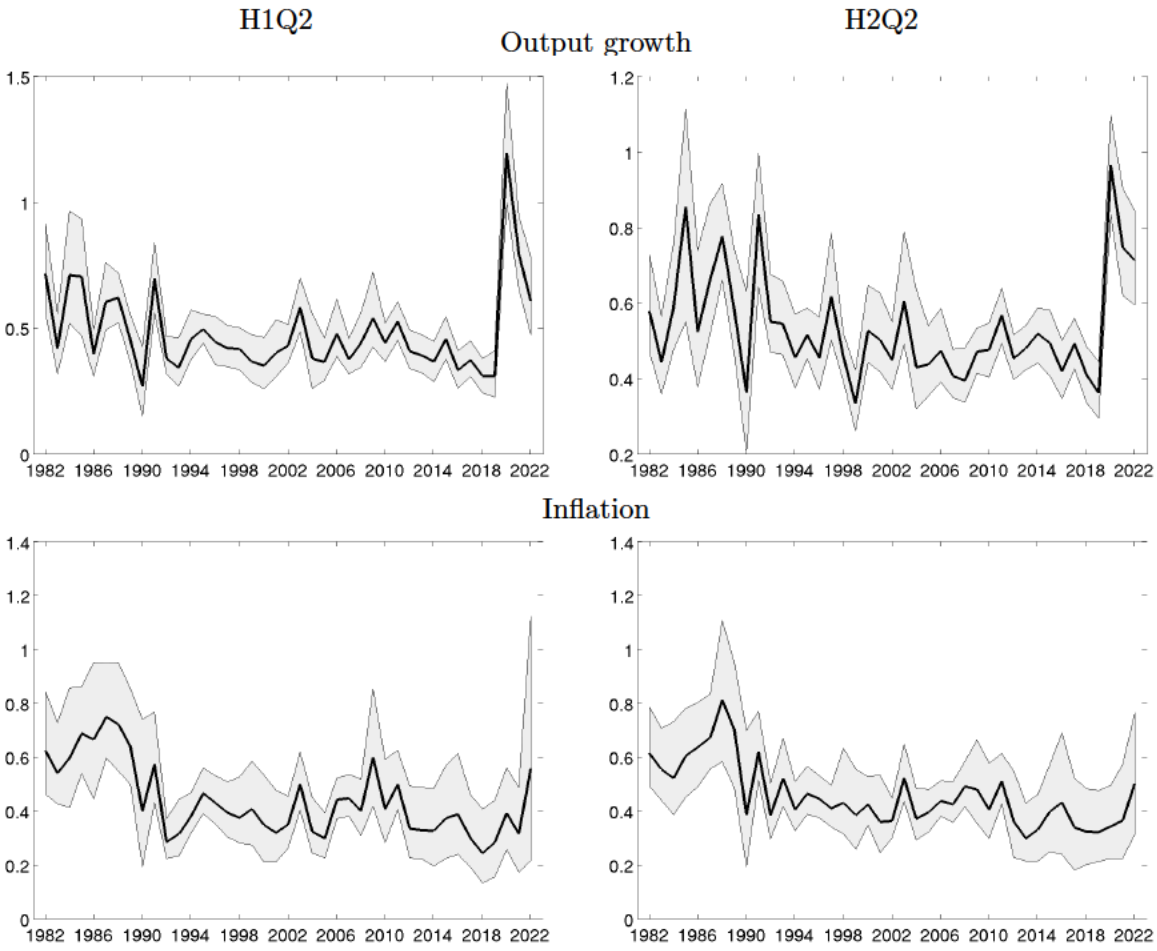
Figure A-50: Subjective uncertainty by individual respondent [Figure 3 under "doubling the noise" prior]



Note: Each panel displays the posterior mean of the standard deviation of the subjective predictive distribution by individual respondent (light gray crosses, connected by thin gray line whenever the respondent appears in consecutive surveys), and the cross-sectional average of the individual standard deviations (dashed black line). Top and bottom panels correspond to output growth and inflation projections. The left and right column correspond to current and next year projections.

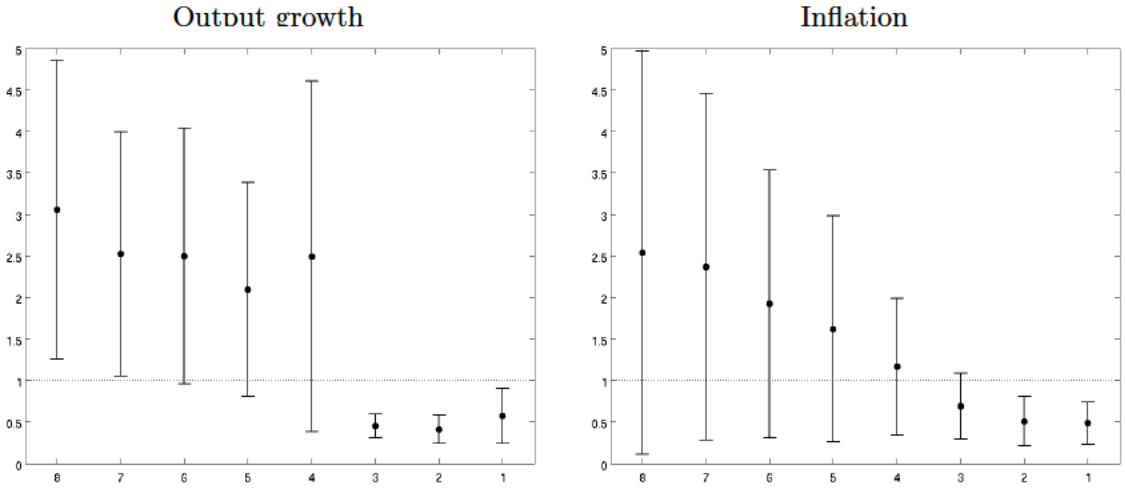
Figure A-51: Cross-sectional standard deviation of individual uncertainty—Q2 survey

[Figure 4 under "doubling the noise" prior]



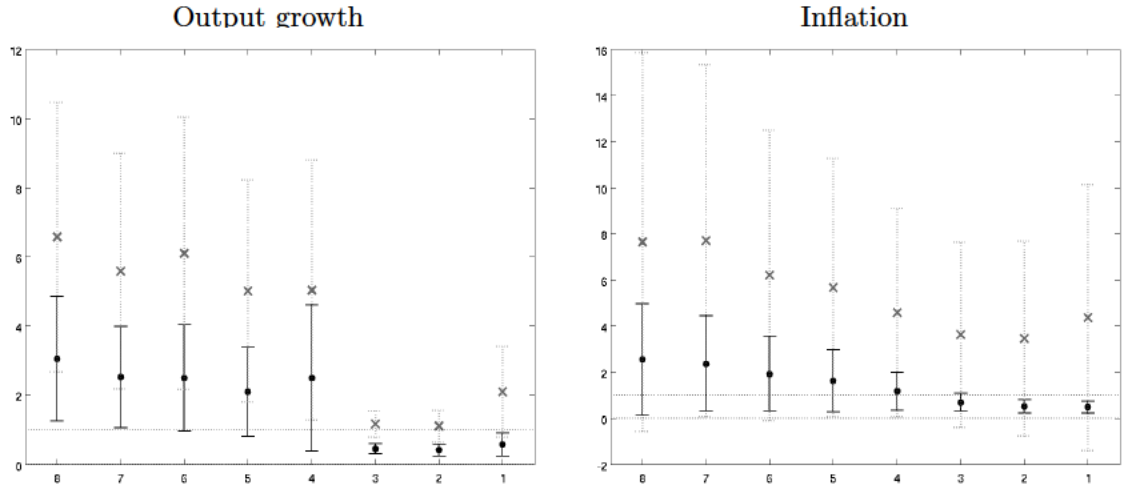
Note: Posterior mean of the cross-sectional standard deviation of the individual standard deviations (solid black line) The shaded areas display the 90 percent posterior credible intervals. Top panels: output growth. Bottom panels: inflation. Left column: current year; right column: next year.

Figure A-52: Do forecasters over or under-estimate uncertainty? A scale test [Figure 5 under "doubling the noise" prior]



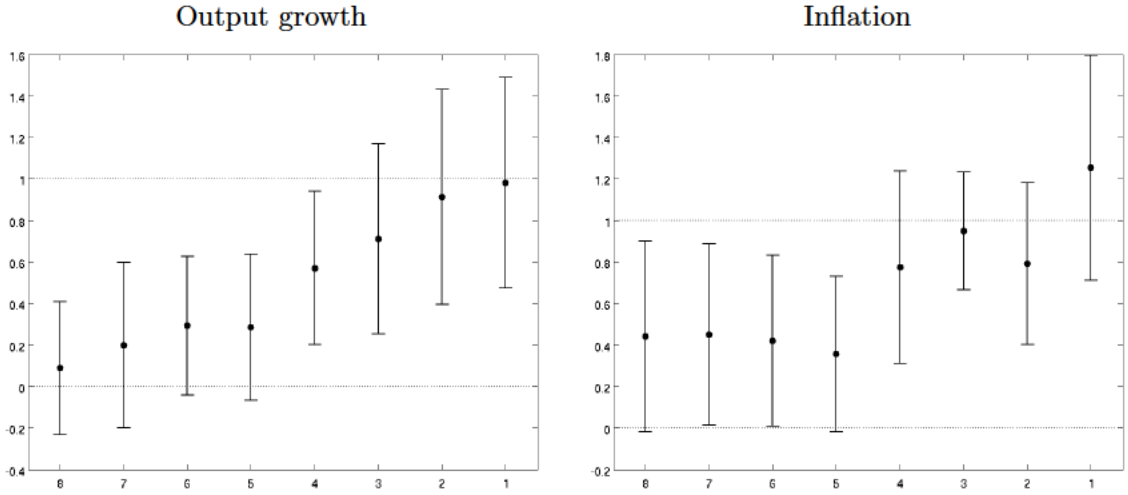
Note: Black dots correspond to OLS estimates of α_q from regression (18) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-53: A scale test—comparison with the generalized beta approach [Figure 6 under "doubling the noise" prior]



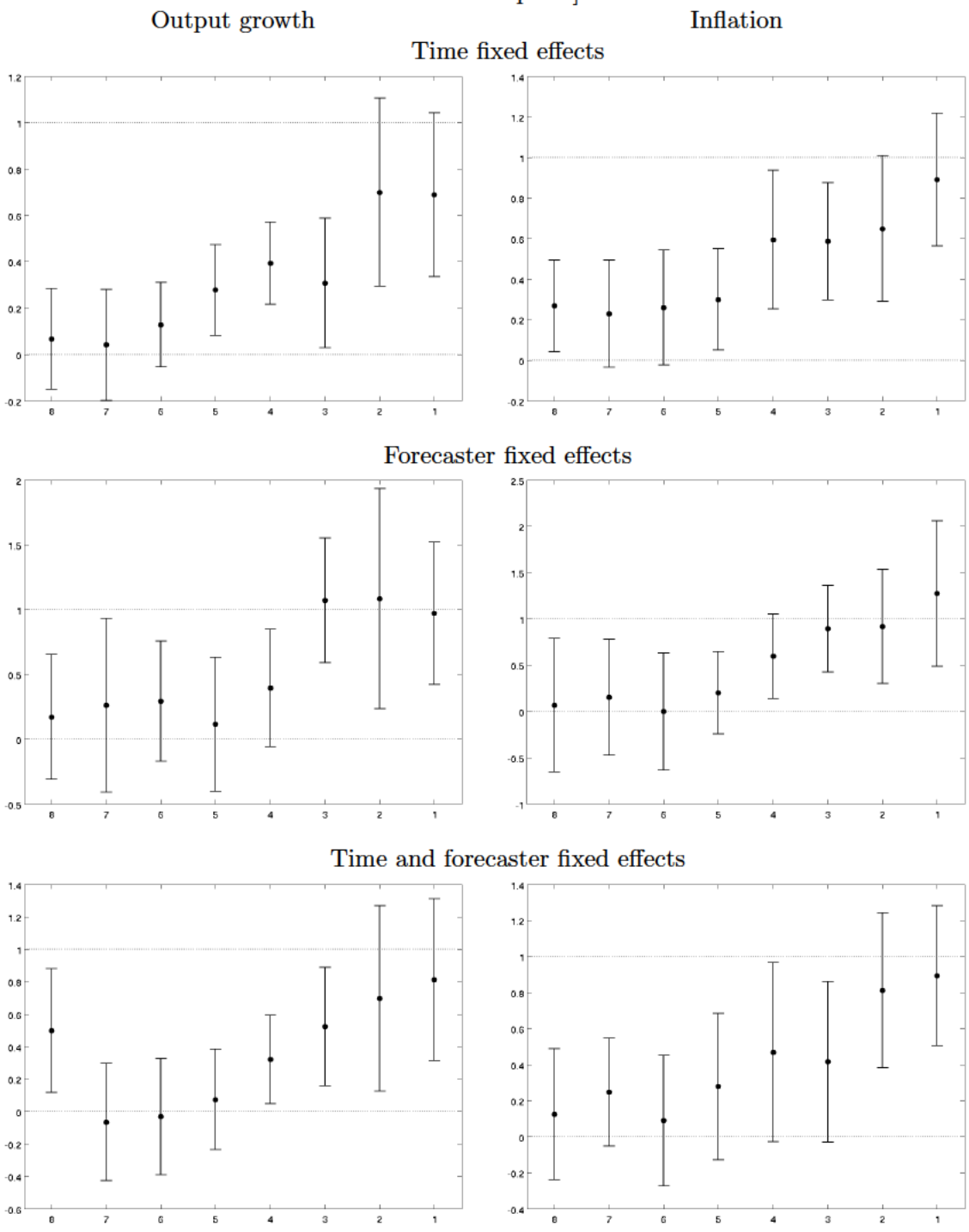
Note: Black dots correspond to OLS estimates of α_q from regression (18) for $q = 8, ..., 1$ using the posterior means for $\mathbb{E}_{t-q,i}[y_t]$ and $\sigma^2_{t|t-q,i}$ from our approach. Gray crosses correspond to OLS estimates when these objects are obtained using the generalized beta approach. Whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-54: Do differences in subjective uncertainty map into differences in forecast accuracy? A variation test [Figure 7 under "doubling the noise" prior]



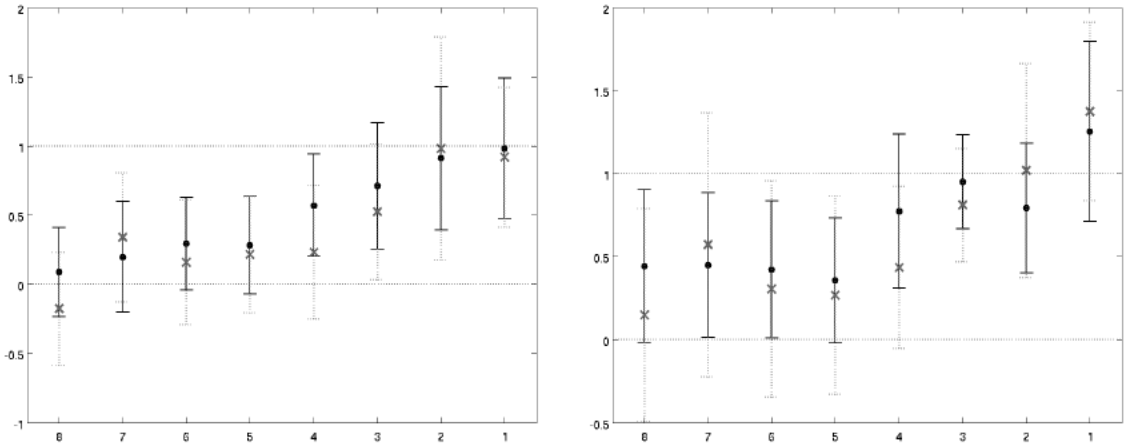
Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-55: A variation test—regressions with fixed effects [Figure 8 under "doubling the noise" prior]



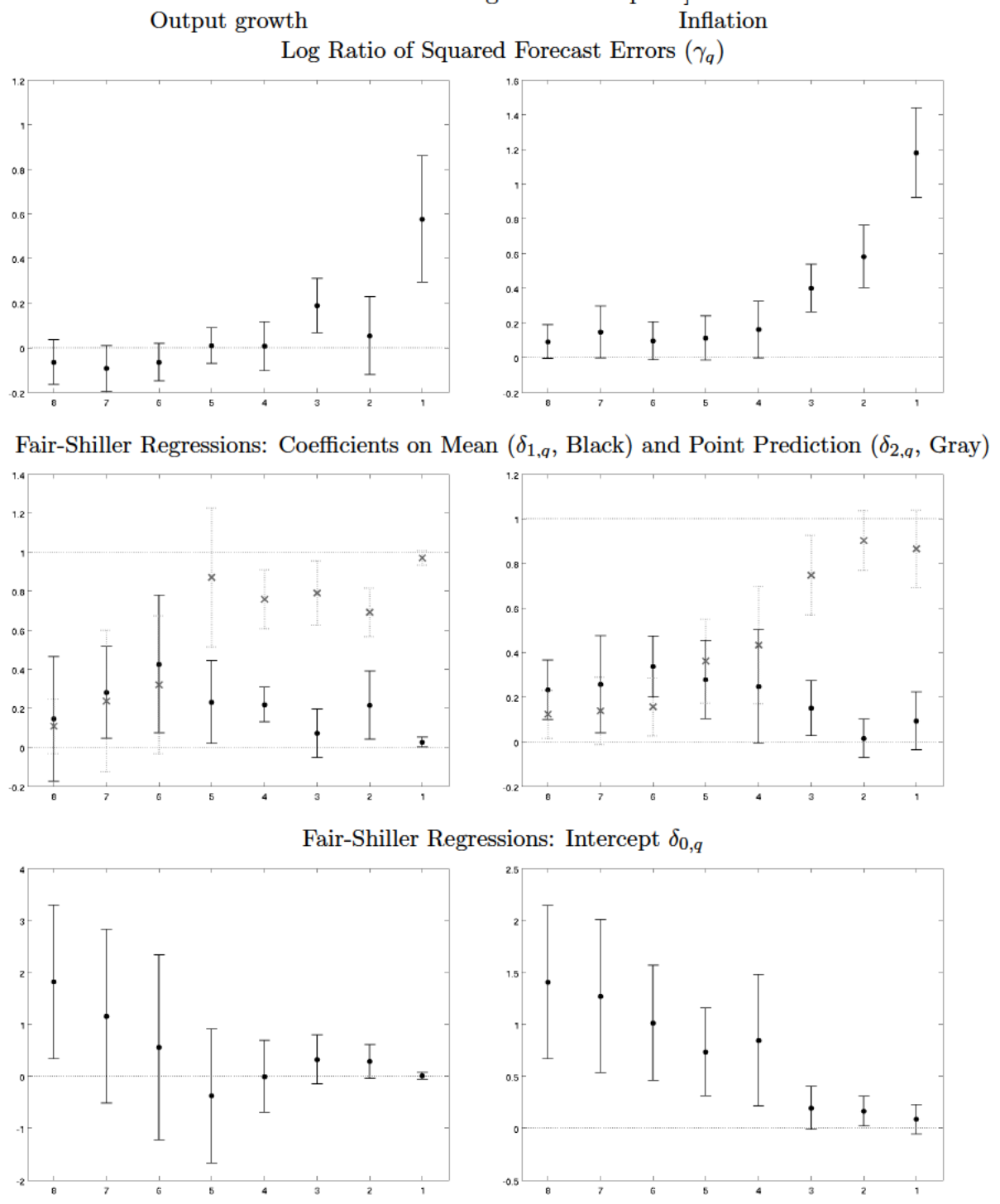
Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) using time (top panels), forecaster fixed effects (middle panels), or both (bottom panels), for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-56: A variation test—accounting for inference uncertainty (baseline vs weighted OLS) [Figure 9 under "doubling the noise" prior]



Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors. Gray crosses correspond to weighted OLS estimates, where the weights are inversely proportional to inference uncertainty as measured by the interquantile range of the posterior distribution of $\sigma_{t|t-q,i}$. Whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-57: A location test: Relative accuracy of mean vs point projections [Figure 10 under "doubling the noise" prior]



Note: Top panel: Black dots correspond to OLS estimates of γ_q from regression (20) for $q = 8, ..., 1$. Middle panel: Black dots and gray crosses correspond to OLS estimates of $\delta_{1,q}$ and $\delta_{2,q}$, respectively, from regression (21) for $q = 8, ..., 1$. Bottom panel: Black dots correspond to OLS estimates of the constant $\delta_{0,q}$ from regression (21) for $q = 8, ..., 1$. In all panels, whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

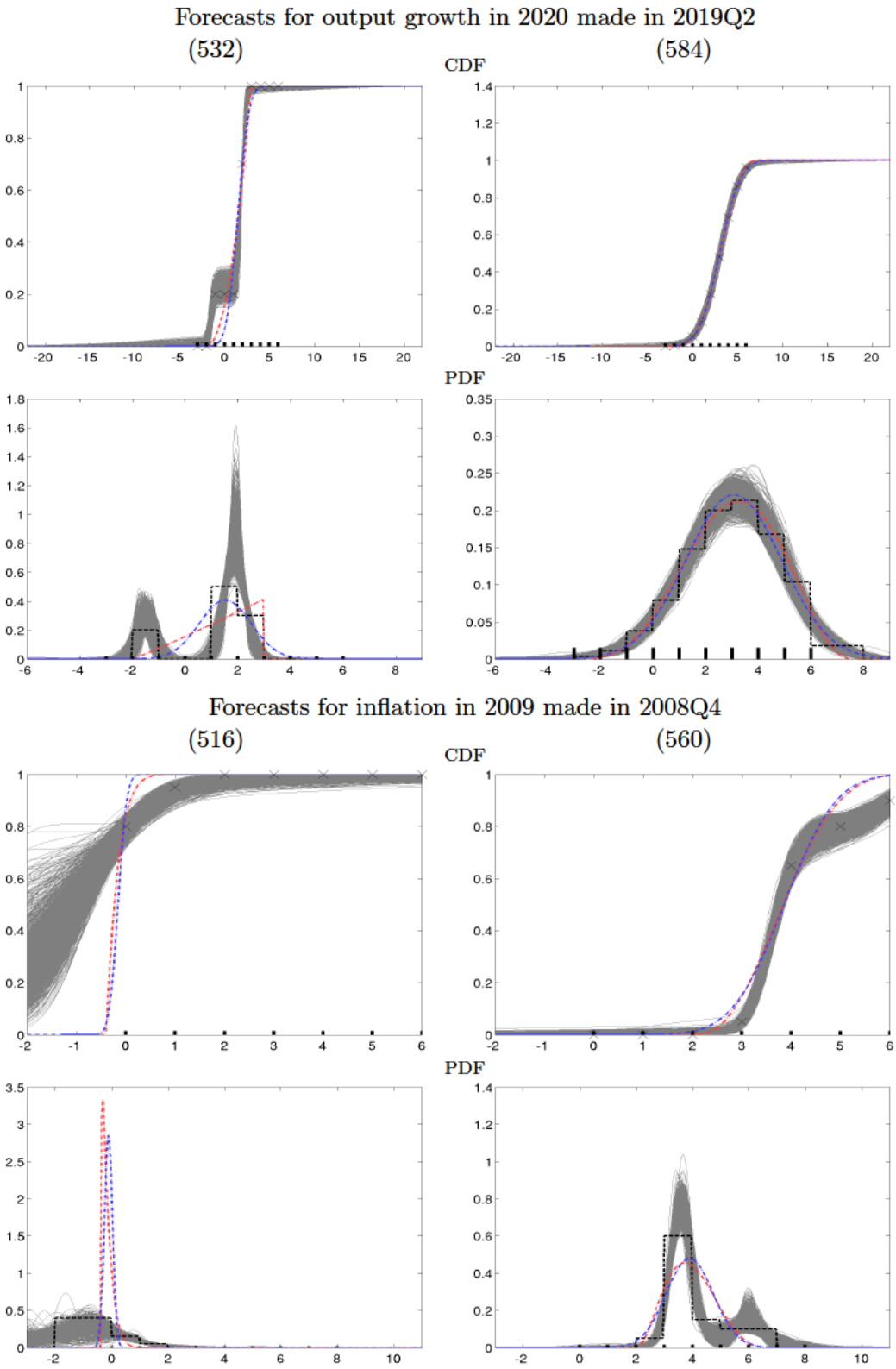
E.C Mixture of Three Normals $F(\cdot)$

Under this robustness check the subjective distribution $F(\cdot|\boldsymbol{\theta})$ is a mixture of three Gaussian distributions, as opposed to two as in (A-26):

$$F(y|\boldsymbol{\theta}) = (1-\omega)\Phi(y|\mu, \sigma_1^2) + \omega(1-\omega_1)\Phi(y|\mu+\mu_\delta, \sigma_2^2) + \omega\omega_1\Phi(y|\mu+\mu_\delta 3, \sigma_3^2). \quad \text{(A-26)}$$

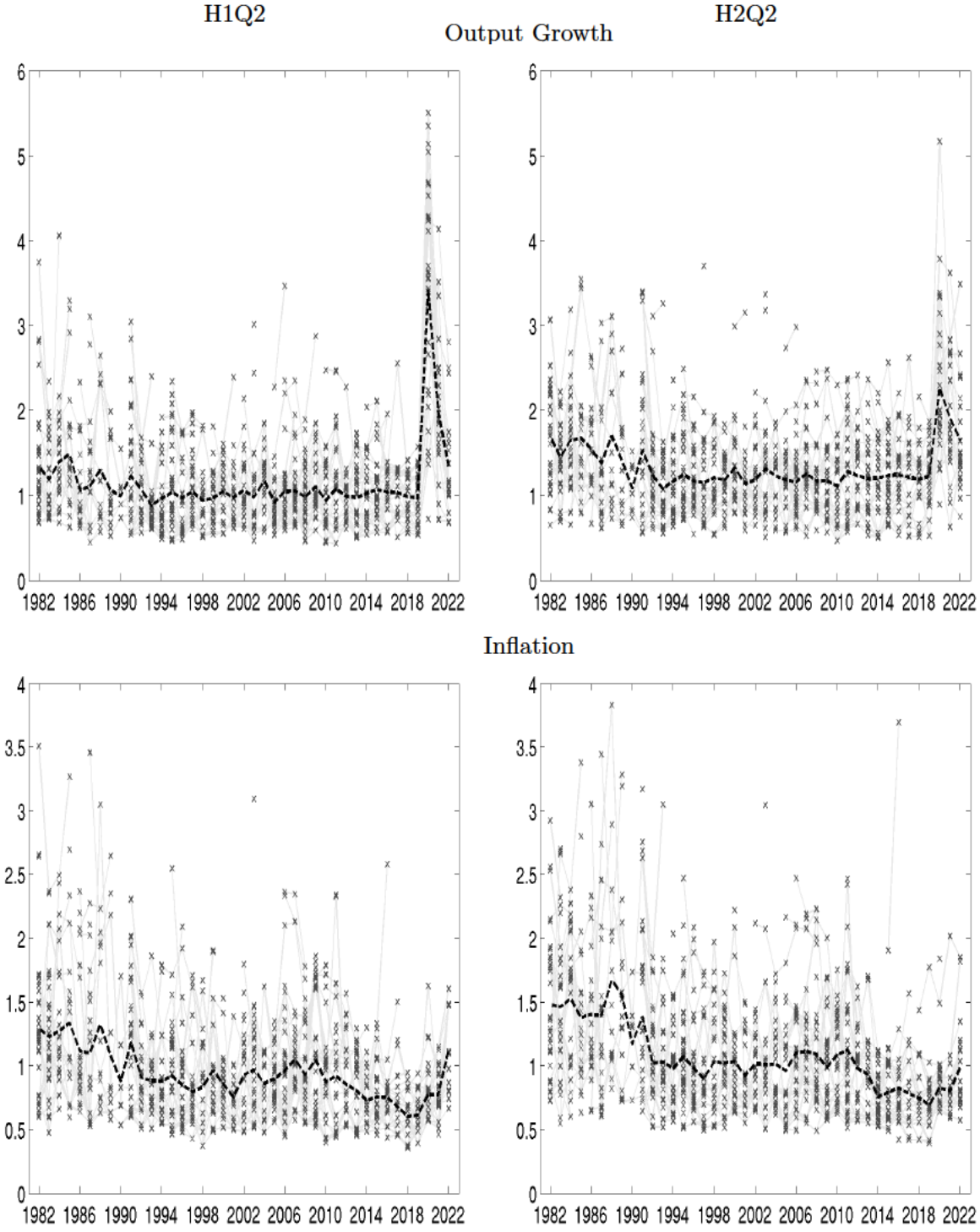
The priors for ω_1, $\mu_\delta 3$, and σ_3 are the same as the priors on ω, μ_δ, and σ_2 described in section II.F.

Figure A-58: Inference using the Bayesian nonparametric approach: CDFs and PDFs for selected examples [Figure 2 under mixture of three Normals $F(\cdot)$]



Note: For each forecaster the top and bottom panels show the subjective CDF and PDF, respectively, estimated using the BNP approach (posterior draws; light gray), as well as the least-squares estimates obtained under the normal (gray, dashed line) and the beta (black, dash-and-dotted line) parametric assumptions. The CDF panels also show the observed cumulated histogram probabilities Z_{ij} $j = 1, \ldots, J$ (crosses), while the PDF panel show the step-wise uniform PDF (gray dashed lines) implied by such probabilities.

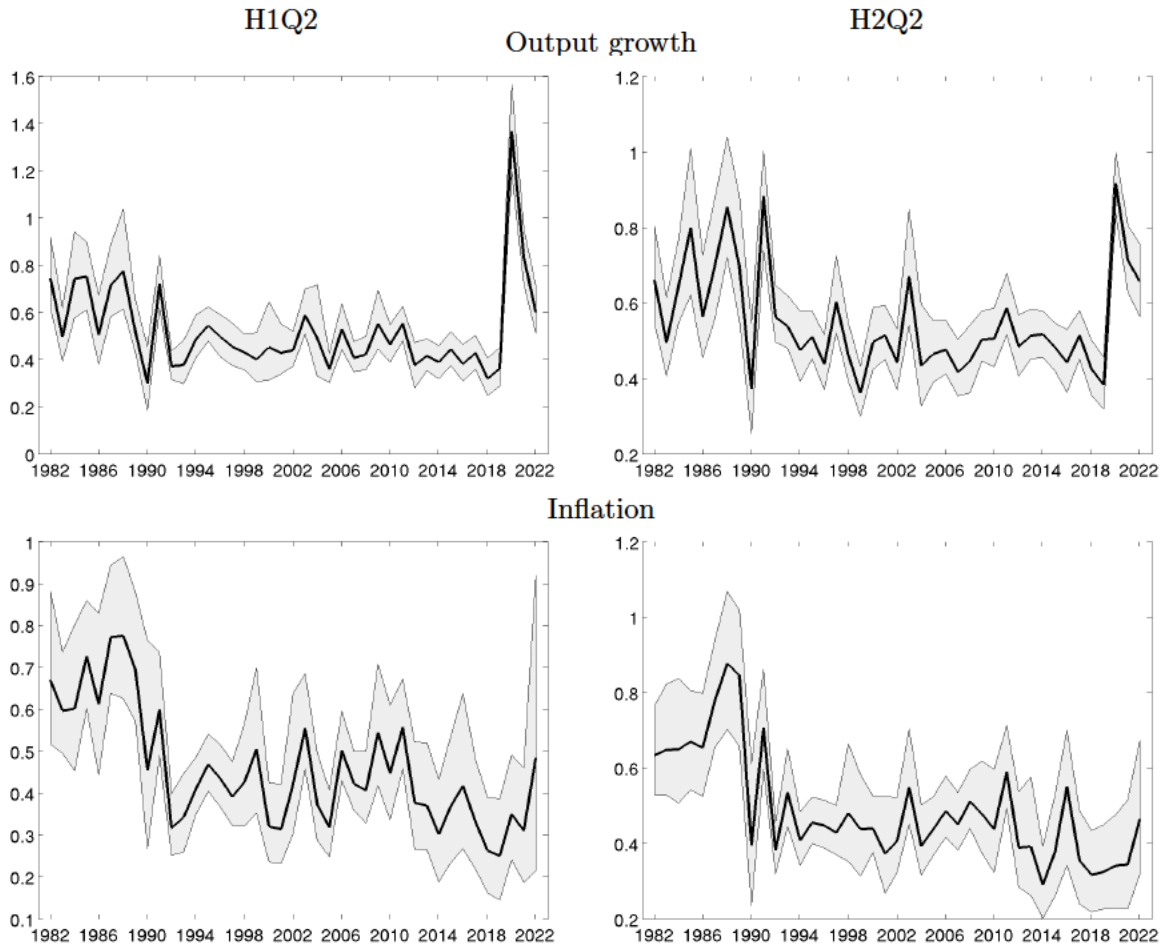
Figure A-59: Subjective uncertainty by individual respondent [Figure 3 under mixture of three Normals $F(\cdot)$]



Note: Each panel displays the posterior mean of the standard deviation of the subjective predictive distribution by individual respondent (light gray crosses, connected by thin gray line whenever the respondent appears in consecutive surveys), and the cross-sectional average of the individual standard deviations (dashed black line). Top and bottom panels correspond to output growth and inflation projections. The left and right column correspond to current and next year projections.

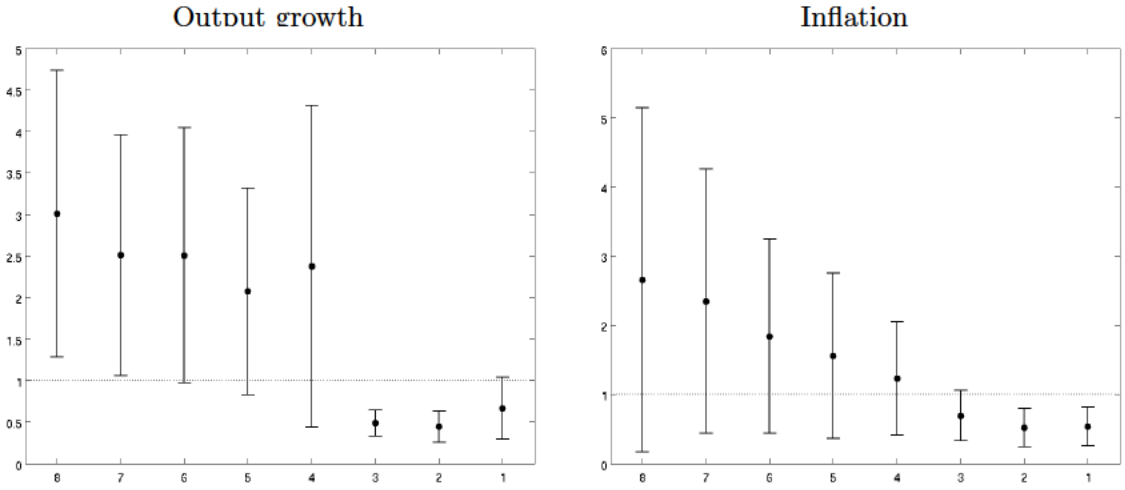
Figure A-60: Cross-sectional standard deviation of individual uncertainty—Q2 survey

[Figure 4 under mixture of three Normals $F(\cdot)$]



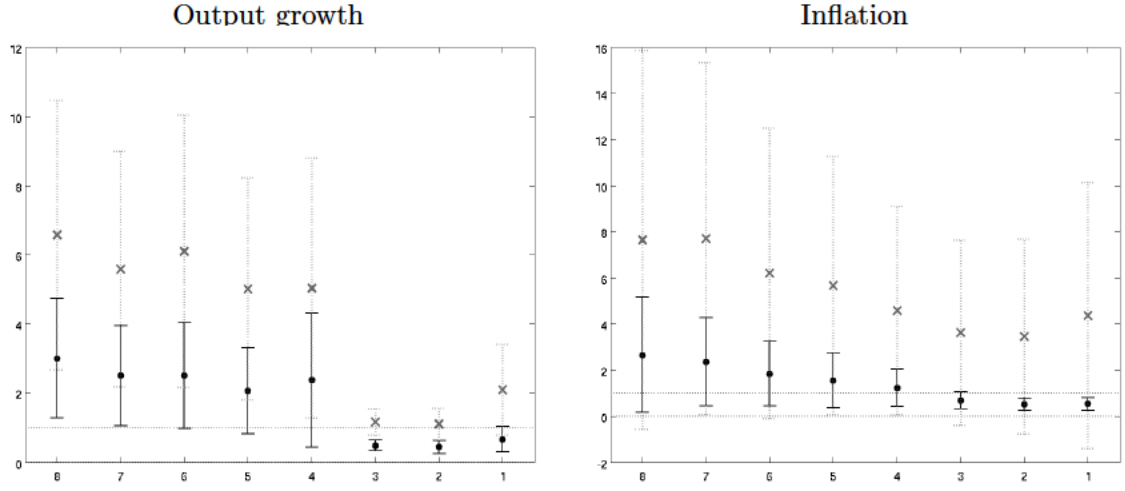
Note: Posterior mean of the cross-sectional standard deviation of the individual standard deviations (solid black line) The shaded areas display the 90 percent posterior credible intervals. Top panels: output growth. Bottom panels: inflation. Left column: current year; right column: next year.

Figure A-61: Do forecasters over or under-estimate uncertainty? A scale test [Figure 5 under mixture of three Normals $F(\cdot)$]



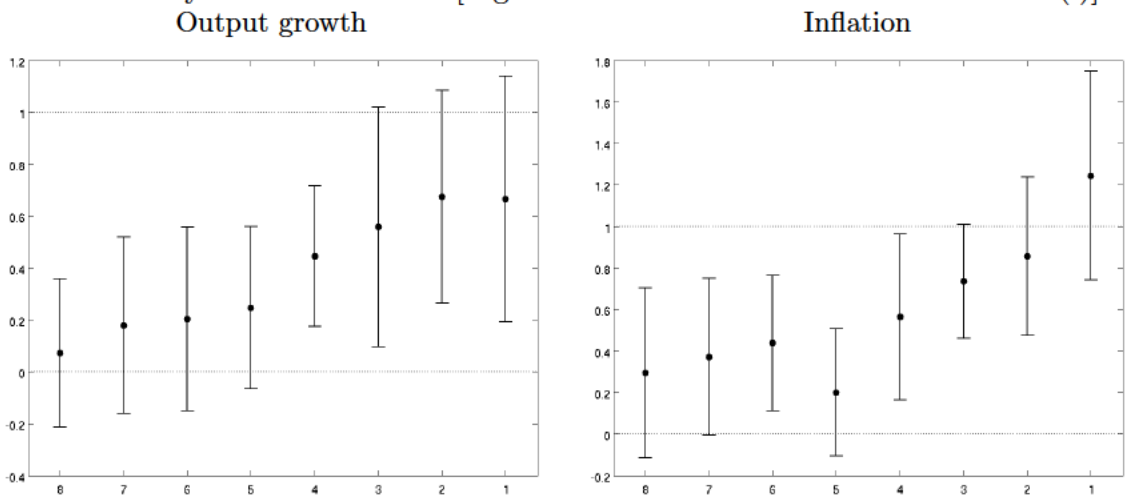
Note: Black dots correspond to OLS estimates of α_q from regression (18) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-62: A scale test—comparison with the generalized beta approach [Figure 6 under mixture of three Normals $F(\cdot)$]



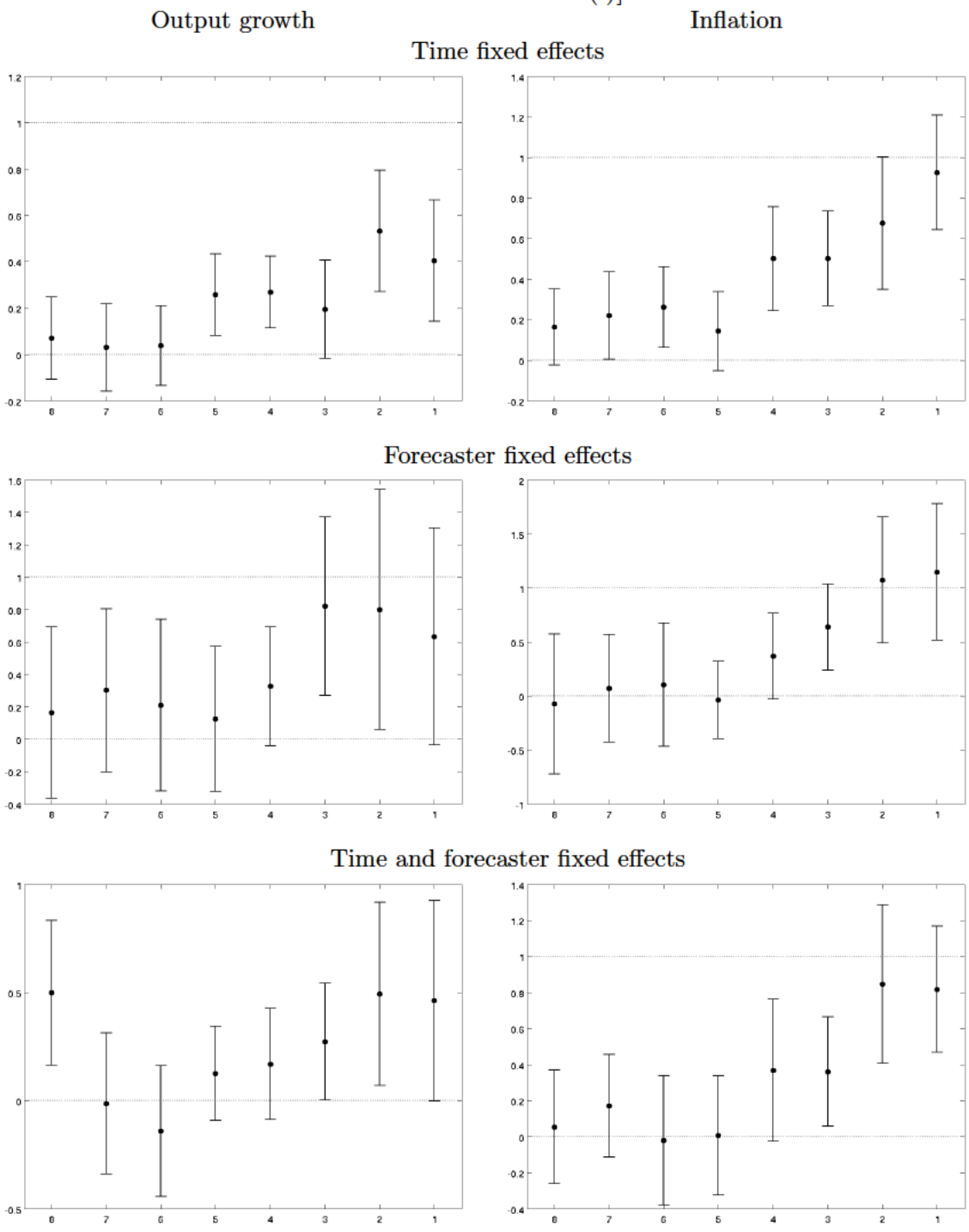
Note: Black dots correspond to OLS estimates of α_q from regression (18) for $q = 8, ..., 1$ using the posterior means for $\mathbb{E}_{t-q,i}[y_t]$ and $\sigma^2_{t|t-q,i}$ from our approach. Gray crosses correspond to OLS estimates when these objects are obtained using the generalized beta approach. Whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-63: Do differences in subjective uncertainty map into differences in forecast accuracy? A variation test [Figure 7 under mixture of three Normals $F(\cdot)$]



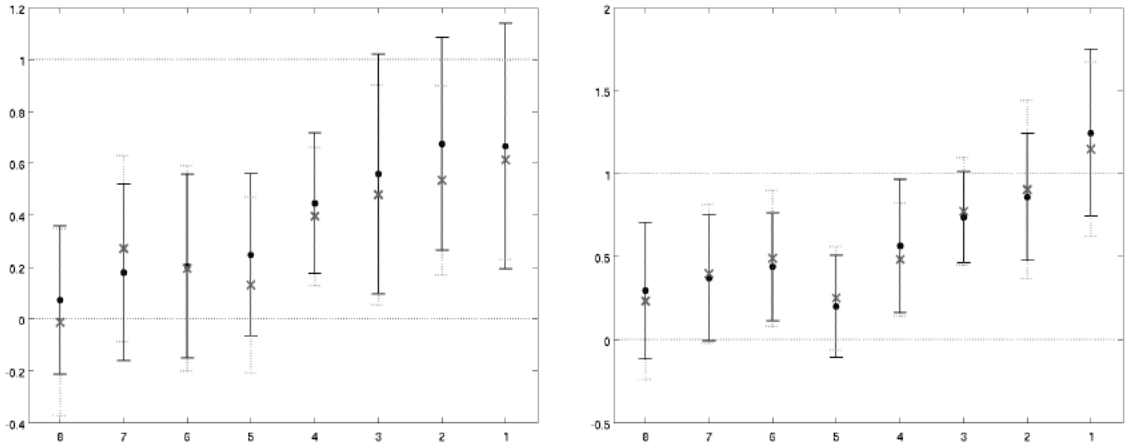
Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-64: A variation test—regressions with fixed effects [Figure 8 under mixture of three Normals $F(\cdot)$]



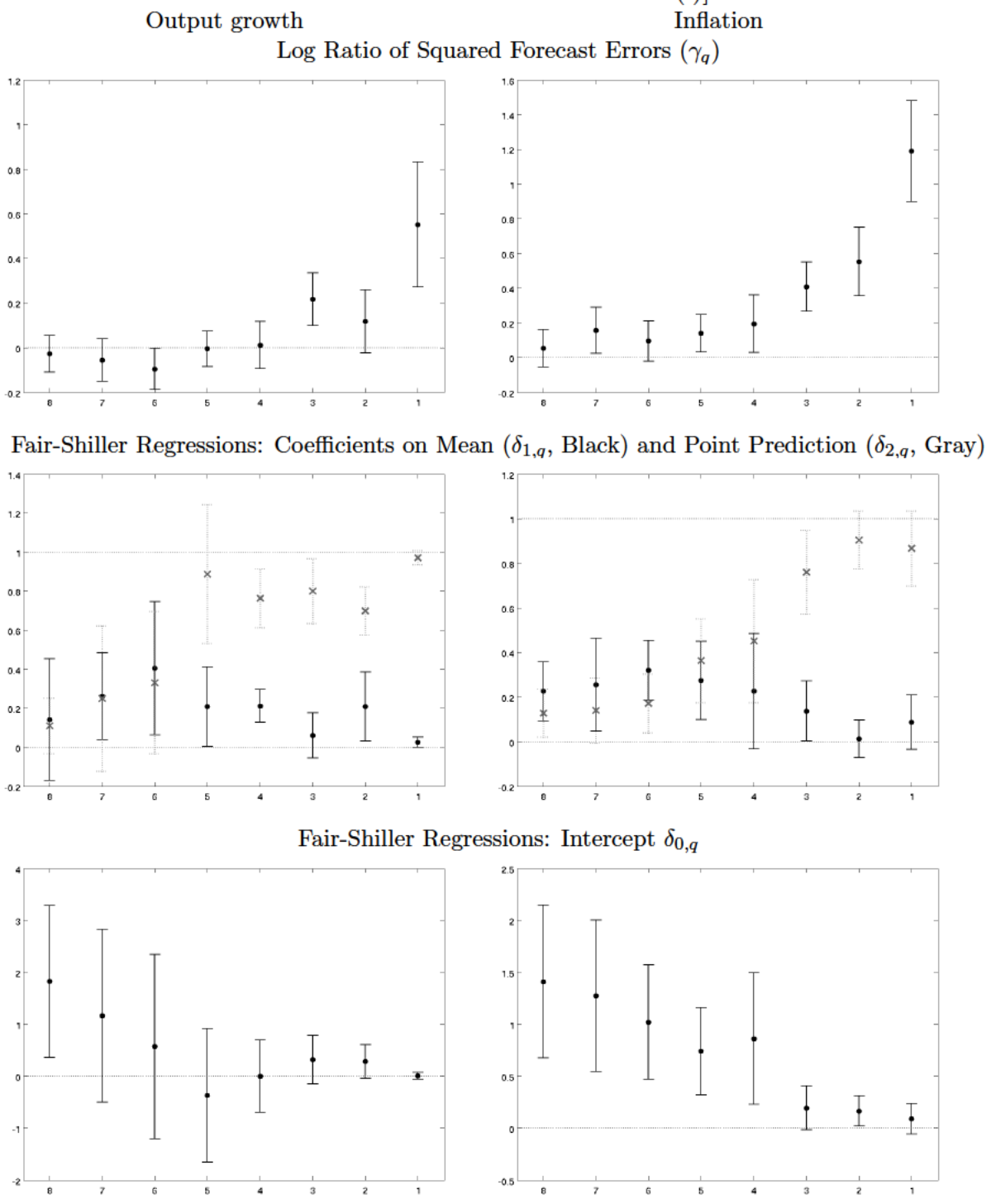
Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) using time (top panels), forecaster fixed effects (middle panels), or both (bottom panels), for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-65: A variation test—accounting for inference uncertainty (baseline vs weighted OLS) [Figure 9 under mixture of three Normals $F(\cdot)$]



Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors. Gray crosses correspond to weighted OLS estimates, where the weights are inversely proportional to inference uncertainty as measured by the interquantile range of the posterior distribution of $\sigma_{t|t-q,i}$. Whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

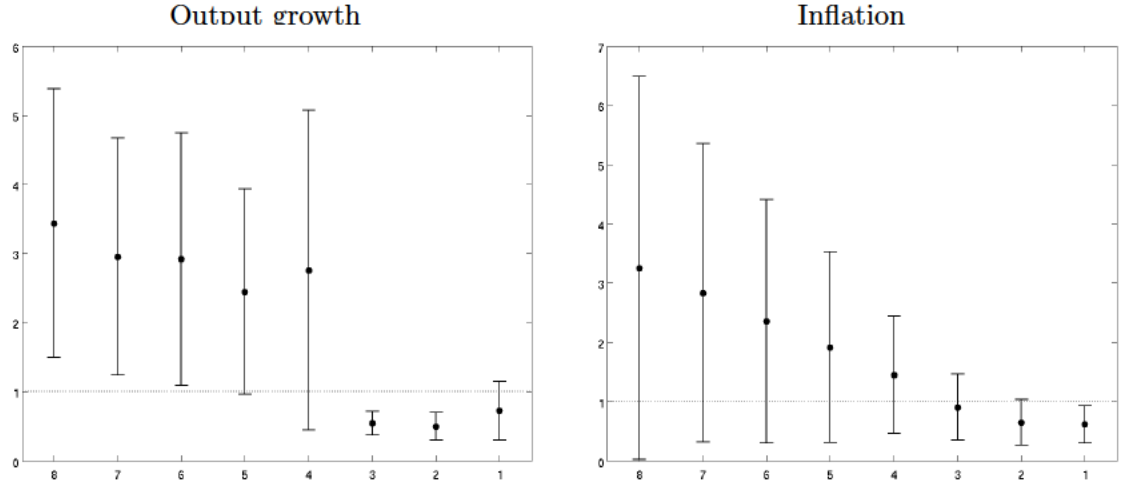
Figure A-66: A location test: Relative accuracy of mean vs point projections [Figure 10 under mixture of three Normals $F(\cdot)$]



Note: Top panel: Black dots correspond to OLS estimates of γ_q from regression (20) for $q = 8, ..., 1$. Middle panel: Black dots and gray crosses correspond to OLS estimates of $\delta_{1,q}$ and $\delta_{2,q}$, respectively, from regression (21) for $q = 8, ..., 1$. Bottom panel: Black dots correspond to OLS estimates of the constant $\delta_{0,q}$ from regression (21) for $q = 8, ..., 1$. In all panels, whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

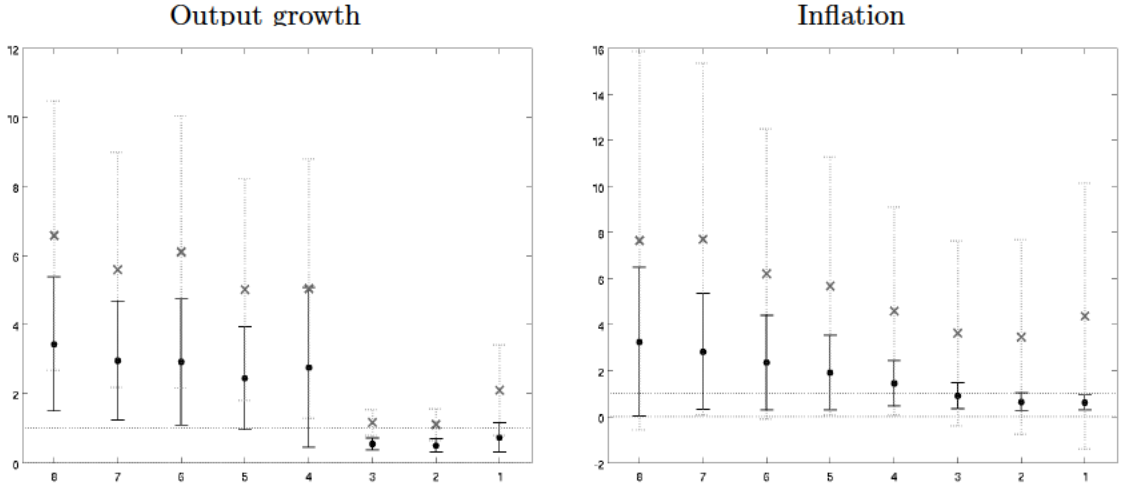
E.D Section III.C results under outlier Winsorization

Figure A-67: Do forecasters over or under-estimate uncertainty? A scale test [Figure 5 under outlier Winsorization (at 2.5%)]



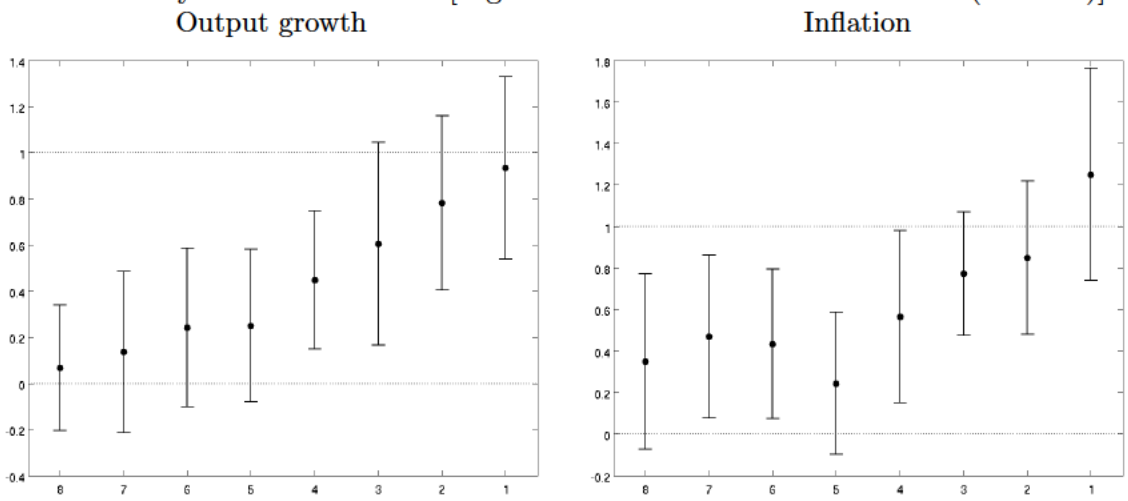
Note: Black dots correspond to OLS estimates of α_q from regression (18) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-68: A scale test—comparison with the generalized beta approach [Figure 6 under outlier Winsorization (at 2.5%)]



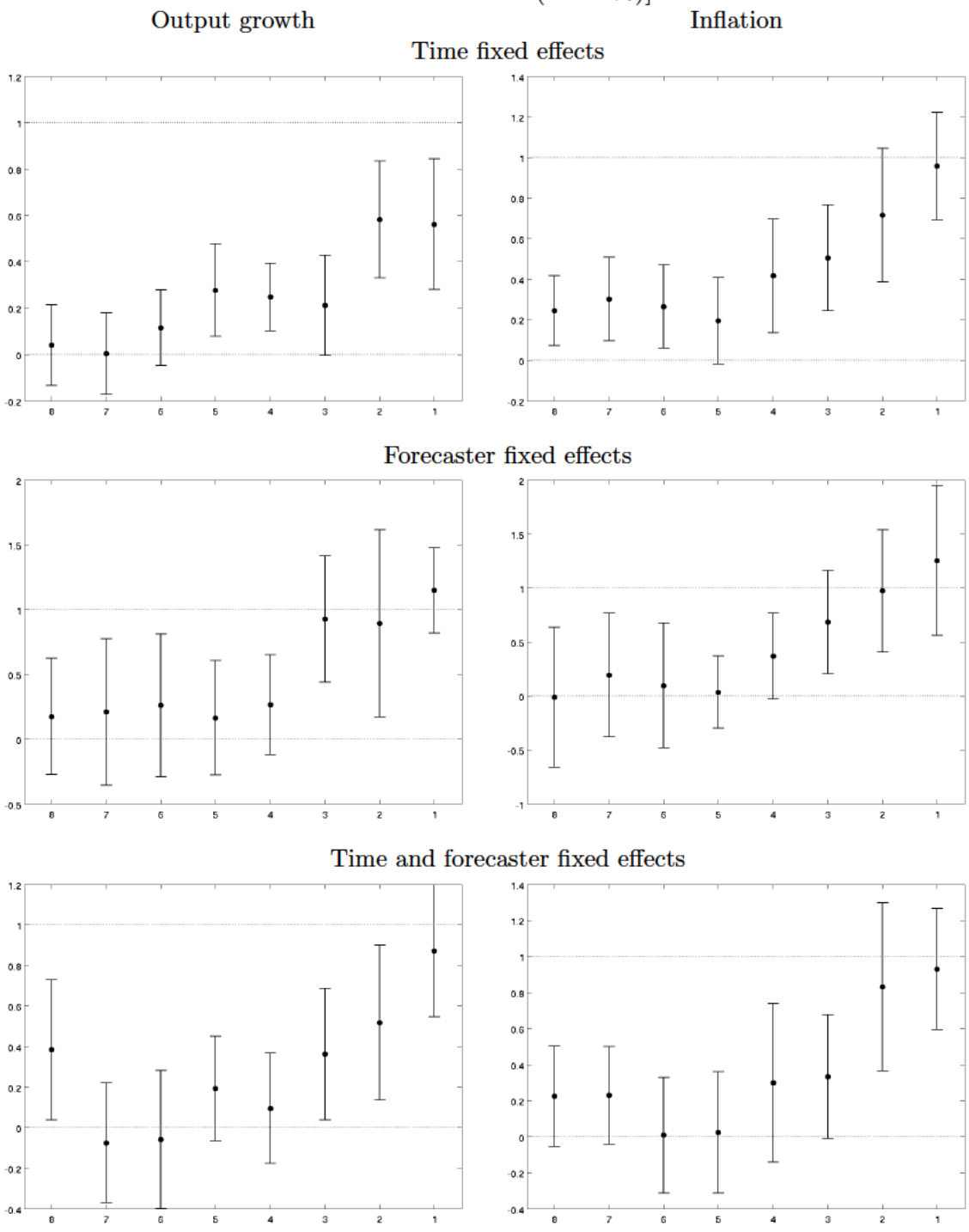
Note: Black dots correspond to OLS estimates of α_q from regression (18) for $q = 8, ..., 1$ using the posterior means for $\mathbb{E}_{t-q,i}[y_t]$ and $\sigma^2_{t|t-q,i}$ from our approach. Gray crosses correspond to OLS estimates when these objects are obtained using the generalized beta approach. Whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-69: Do differences in subjective uncertainty map into differences in forecast accuracy? A variation test [Figure 7 under outlier Winsorization (at 2.5%)]



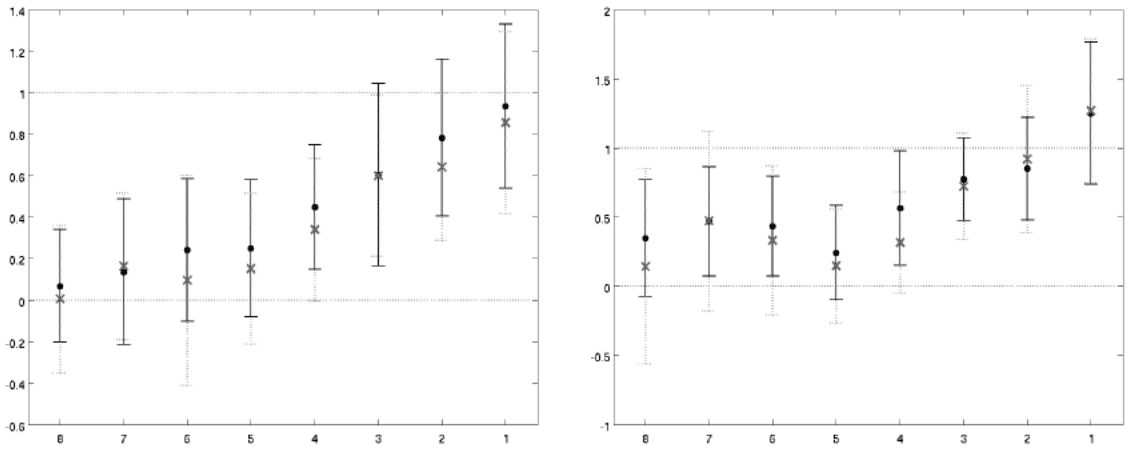
Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-70: A variation test—regressions with fixed effects [Figure 8 under outlier Winsorization (at 2.5%)]



Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) using time (top panels), forecaster fixed effects (middle panels), or both (bottom panels), for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

Figure A-71: A variation test—accounting for inference uncertainty (baseline vs weighted OLS) [Figure 9 under outlier Winsorization (at 2.5%)]

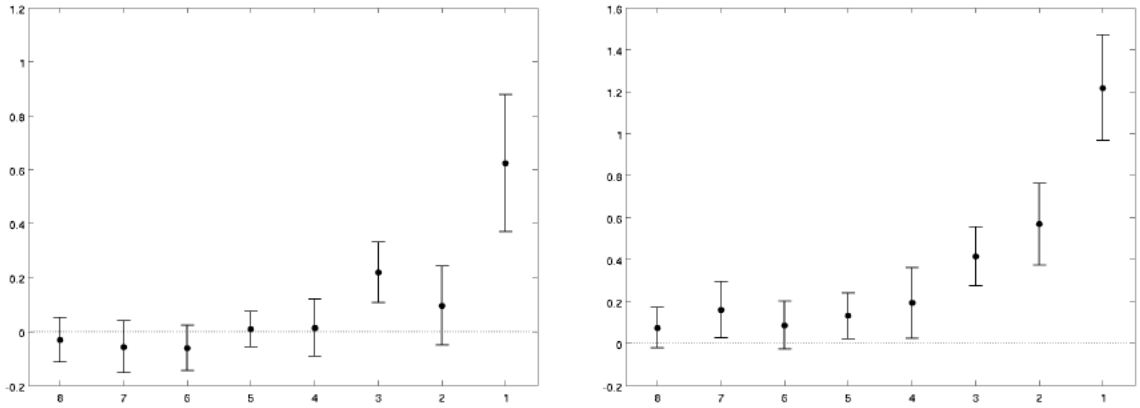


Note: Black dots correspond to OLS estimates of $\beta_{1,q}$ from regression (19) for $q = 8, ..., 1$. Solid black whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors. Gray crosses correspond to weighted OLS estimates, where the weights are inversely proportional to inference uncertainty as measured by the interquantile range of the posterior distribution of $\sigma_{t|t-q,i}$. Whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.

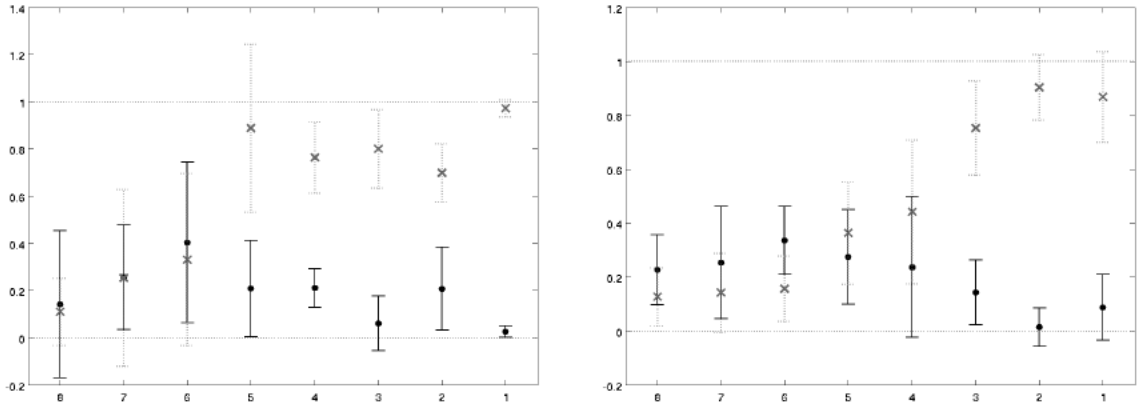
Figure A-72: A location test: Relative accuracy of mean vs point projections [Figure 10 under outlier Winsorization (at 2.5%)]

Output growth Inflation

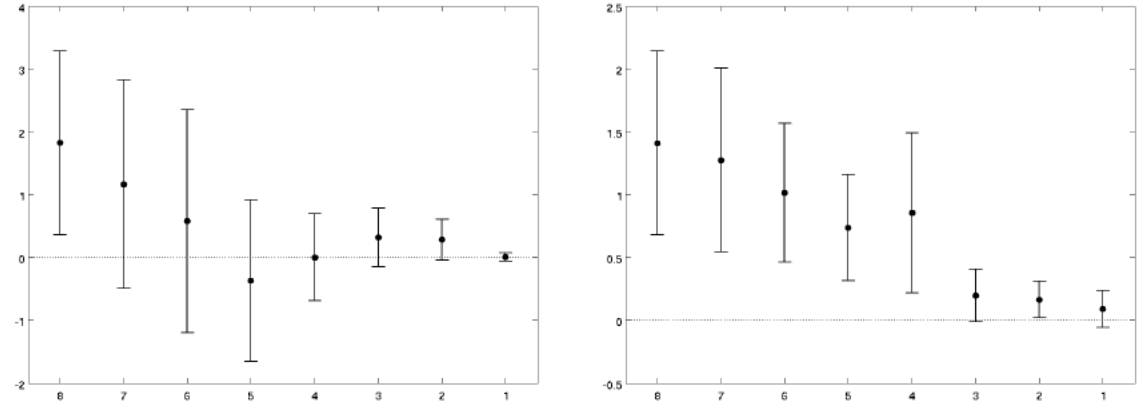
Log Ratio of Squared Forecast Errors (γ_q)



Fair-Shiller Regressions: Coefficients on Mean ($\delta_{1,q}$, Black) and Point Prediction ($\delta_{2,q}$, Gray)



Fair-Shiller Regressions: Intercept $\delta_{0,q}$



Note: Top panel: Black dots correspond to OLS estimates of γ_q from regression (20) for $q = 8, ..., 1$. Middle panel: Black dots and gray crosses correspond to OLS estimates of $\delta_{1,q}$ and $\delta_{2,q}$, respectively, from regression (21) for $q = 8, ..., 1$. Bottom panel: Black dots correspond to OLS estimates of the constant $\delta_{0,q}$ from regression (21) for $q = 8, ..., 1$. In all panels, whiskers indicate 90 percent posterior coverage intervals based on Driscoll-Kraay standard errors.