The Conquest of Inflation Credibility in the U.S.

A Bayesian Approach for Inference on Probabilistic Surveys

Marco Del Negro*
Federal Reserve Bank of New York

Roberto Casarin
University Ca' Foscari of Venice

Francesco Ravazzolo
University of Bozen/Bolzano

February 20, 2018

INCOMPLETE DRAFT – PLEASE DO NOT CIRCULATE, DISTRIBUTE, OR POST

Abstract

We propose a non-parametric Bayesian approach to the estimation of forecast densities in probabilistic surveys. We use it to study the evolution of the subjective forecast distribution for the U.S. Survey of Professional Forecasters over the past forty years, focusing especially on second moments. We show that the variance of aggregate forecast distribution fell substantially from the eighties to the nineties (the "conquest"), and fell again after the Fed announced its long term inflation goal. We also show that disagreement (heterogeneity in the mean forecasts) plays a minor role, but that heterogeneity in uncertainty is very large. The "conquest" amounted to convincing high-uncertainty forecasters that inflation is under control. We also find that only a fringe of forecasters place any significant probability of the possibility of a return to the seventies. The likelihood of deflation in the aftermath of the Great Recession was significant (almost ten percent for the average forecaster) but declined to one percent or less for most forecasters thereafter.

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

KEY WORDS: Bayesian inference, Bayesian Nonparametric, Survey of Professional Forecasters, Inflation Credibility.

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

I Introduction

The pioneering work of Manski (2004) made economists appreciate the advantages of probabilistic surveys relative to surveys that simply ask respondents for their point projections.[1] First of all, with point forecast we are never sure of what we are getting, since responses depend on the forecaster's loss function: they could be reporting the mean, the median, or some other quantile of their distribution. More importantly, probabilistic surveys provide a wealth of information that is not included in point projections. 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, the median, the variance, the skewness, the interquantile range, *et cetera*.

Except that 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 pre-specified 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 most quantities of interest, standard practice consists in making a parametric assumption concerning the forecast distribution and computing its parameters by minimizing the distance between the observed CDF points and those implied by the assumed distribution, following the approach in Engelberg et al. (2009).[2] This is clearly an inference problem – we do not know the parameters of this distribution nor, for that matter, its parametric form. Yet the literature has so far sidestepped this issue, at least to our knowledge, and has reported objects of interest as if they were devoid of uncertainty.

We address this problem by proposing a non-parametric approach for the estimation of the survey respondents' forecast densities. Specifically, we model these unknown distributions without resorting to strong parametric assumptions via a non-parametric Bayesian model with Dirichlet process priors. The non-parametric dimension allows for a flexibility that the conventional approach – which amounts to using a generalized Beta distribution,

[1] Indeed the most recent surveys, such as the Federal Reserve Bank of New York Survey of Consumer Expectations, relies heavily on probabilistic questions (see Potter (2016)).

[2] For very few quantities of interest, such as the median, one can compute non-parametric bounds as in Engelberg et al. (2009). Such bounds can also be computed for the mean if one is willing to "close" the open bins.

again following Engelberg et al. (2009) – does not entirely afford.[3] In economics and finance, the Bayesian non-parametric approach so far has applied to the analysis of treatment effects (Chib and Hamilton, 2002), autoregressive panel data (Hirano, 2002), job search (Koop, 2003), stochastic production frontiers models (Griffin and Steel, 2004), unemployment duration (Burda et al., 2015), and models of stochastic volatility in asset returns (Griffin, 2011, and Jensen and Maheu, 2010) (see Griffin et al., 2011, for a recent survey of Bayesian non-parametric and semiparametric methods, and their applications in economics). Outside of economics, these methods are widely used in biostatistics (Mitra and Müller, 2015), machine learning (Blei et al., 2010, Hannah et al., 2011), and psychology (Griffiths and Tenenbaum, 2006).

We use this approach to to study the evolution of the subjective forecast distribution for the U.S. Survey of Professional Forecasters over the past forty years, focusing especially on second moments. We show that the variance of aggregate forecast distribution fell substantially from the eighties to the nineties (the "conquest"), and fell again after the Fed announced its long term inflation goal. We also show that

1. The fall in disagreement plays a truly minor role. It is relevant only in the eighties and even then it is due to one or two outliers.

2. Almost all of the decline in the variance of the aggregate distribution is due to a decline in average uncertainty. This is a result robust to omitting outliers.

3. The decline in uncertainty is not at all homogeneous across forecasters. In fact, we show that for about half of the forecasters inflation uncertainty was very low to start with, even in the eighties. The "conquest of inflation credibility" therefore amounts to "high-uncertainty" forecasters becoming convinced that inflation was under control.

Next, we study the probability of a return to the inflation levels of the seventies, as viewed from the perspective of the SPF forecasters. We think this is an interesting question because it provides indirect evidence on the empirical appeal of recent DSGE models with regime switching (e.g. Bianchi and Ilut (2015), Bianchi and Melosi (2014); see Sims and Zha

[3] The generalized Beta adopted by Engelberg et al. (2009) consists in a standard Beta distribution, except that its domain is not constrained to be between zero and one. This distribution does not accommodate multi-modality, and forces the researcher to arbitrarily close the open bins. Dominitz and Manski (1997) use a lognormal distribution.

(2006) for a VAR version). In these models one of the regime can indeed be characterized as a "back to the seventies" regime. Do forecasters truly entertain such a possibility?

Finally, we study the likelihood of deflation, again as viewed from the perspective of the SPF forecasters.

II Bayesian Inference for Probabilistic Surveys

For each forecaster $i = 1, ..., n$ the available data consists of a vector of probabilities $z_i = (z_{i,1}, \ldots, z_{i,J})$, with $z_{i,j} \geq 0$ and $\sum_{j=1}^{J} z_{i,j} = 1$, measuring the predictive likelihood that continuous variable y (in this application, inflation) falls within the respective bin. The J bins are mutually exclusive and contiguous, and generally cover the entire real line. We denote with $\bar{y}_1 < \ldots < \bar{y}_J$ the set of bin upper bounds, with $\bar{y}_J$ possibly being ∞. We begin by modeling a single cross-section of forecasts, and leave the modeling of the panel to future work. The remainder of the section is structured as follows. The next subsection shows that the non-parametric model implies that the $\mathbf{z}_i$'s distribution is characterized by infinite mixture representation. Section II.B describes the mixture components, and section II.C provides a brief intuition for the Gibbs sampler. In the remainder of the paper we will use the notation $x_{1:k}$ to denote the sequence $\{x_1, .., x_k\}$ for a generic variable x.

II.A The Bayesian Non-Parametric Model

We express our uncertainty over the distribution generating the $\mathbf{z}_i$'s using a Bayesian non-parametric model. Specifically, each $\mathbf{z}_i$ is independently distributed according to

$$h(\mathbf{z}|G) = \int h(\mathbf{z}|\boldsymbol{\theta})G(d\boldsymbol{\theta}) \tag{1}$$

where $h(\cdot|\boldsymbol{\theta})$ is the kernel density, which depends on the parameter vector $\boldsymbol{\theta}$, and which we discuss in detail below, and G is a random probability measure over the parameter space $\boldsymbol{\theta}$. This random probability measure is given by the Dirichlet process (see Ferguson (1973)) $\mathcal{DP}(\psi, G_0)$ with base measure G_0 (that is, $E[G(\boldsymbol{\theta})] = G_0(\boldsymbol{\theta})$) and precision parameter ψ, which measures the concentration of G around G_0.[4]

[4] See Ghosh and Ramamoorthi (2003) for an introduction to Dirichlet process priors and Hjort et al. (2010) for a review on the state-of-the-art practice of Bayesian non-parametrics. Norets and Pelenis (2014)

When $\psi \to \infty$ 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. Outside of this limiting case, the discreteness of the Dirichlet process generates a priori dependency among the $\boldsymbol{\theta}_i$'s via the formation of "clusters". In fact, as shown in Escobar and West (1995), the prior distribution of $\boldsymbol{\theta}_n$ conditional on $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{n-1})$ is given by

$$\boldsymbol{\theta}_n|\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{n-1} \sim \frac{\psi}{\psi + n - 1} G_0(\boldsymbol{\theta}_n) + \frac{1}{\psi + n - 1} \sum_{i=1}^{n-1} \delta_{\boldsymbol{\theta}_i}(d\boldsymbol{\theta}). \tag{2}$$

With probability $\frac{\psi}{\psi + n - 1}$ the new draw $\boldsymbol{\theta}_n$ is generated from G_0, but it is otherwise equal to one of the previous $n-1$ draws. In fact, the n forecasters' distribution can be characterized using N different "clusters", where N is a random variable with prior mean $E[N] \approx \psi log(\frac{\psi + n}{\psi})$.

Following Sethuraman (1994) G can be represented as a discrete random measure

$$G(d\boldsymbol{\theta}) = \sum_{k=1}^{\infty} w_k \delta_{\boldsymbol{\theta}_k}(d\boldsymbol{\theta}) \tag{3}$$

where $\delta_{\boldsymbol{\theta}_k}(\cdot)$ denotes the unit point mass at $\boldsymbol{\theta}_k$, with random weights w_k generated by the stick-breaking construction

$$w_k = v_k \prod_{l=1}^{k-1} (1 - v_l) \tag{4}$$

where the so-called "atoms" $\boldsymbol{\theta}_k$ are i.i.d. random variables from the base measure G_0 and where the stick-breaking components v_l are i.i.d. random variables from a Beta distribution $\mathcal{B}e(1, \psi)$. In this infinite mixture representation, the clusters are captured by the weights. Sethuraman (1994)'s constructive representation, in addition to being computationally convenient as we will see below, implies that our model has the infinite mixture representation

$$h(\mathbf{z}|G) = \sum_{k=1}^{\infty} w_k h(\mathbf{z}|\boldsymbol{\theta}_k) \tag{5}$$

where the weights $w_{i,k}$ come from the same prior distribution (4) for all forecasters.

study posterior consistency in Bayesian non-parametric inference. Griffin and Steel (2011) and Bassetti et al. (2014), among others, introduce dependent Dirichlet process priors over time and in the cross-section, respectively.

II.B The Mixture Components

We view our data – the reported probabilities associated with each bin, $\mathbf{z}$'s – as noise-ridden measurements of the forecasters' predictive distribution (since the mixture components are defined in the same way for all forecasters we omit the subscript i in this section). That is, we assume that each forecaster has a true subjective probability distribution for y, whose cdf is called $F(y|\boldsymbol{\varphi})$, which is associated with a vector of probabilities over the J bins:

$$\nu_j(\boldsymbol{\varphi}) = F(y_j|\boldsymbol{\varphi}) - F(y_{j-1}|\boldsymbol{\varphi}),\ j = 1, \ldots, J, \tag{6}$$

where $y_0 = -\infty$ and, of course, $\nu_j \geq 0$ and $\sum_{j=1}^{J} \nu_j = 1$ since $F(\cdot)$ is a proper probability distribution. In our application we use a Gaussian cdf, i.e. $F(y, \boldsymbol{\varphi}) = \Phi(y|\mu, \sigma^2)$, so that $\boldsymbol{\varphi} = (\mu, \sigma^2)'$.

The reported vector of probabilities $\mathbf{z}$ can potentially differ from $\boldsymbol{\nu}$ because of "noise", which captures approximations, rounding off – for instance, forecasters may report zero when the underlying probability associated to each bin is small – or to actual mistakes in reporting. We need to respect the constraint that the z_j's still need to sum up to one when modeling the noise. We therefore use (again) the Dirichlet distribution, which is one of the most used distributions for compositional data (that is, data in the simplex; see Pawlowsky-Glahn et al., 2015, for an introduction to compositional data modeling). A drawback of the Dirichlet distribution is that its pdf is null for $\mathbf{z}$'s that have some elements equal to zero, when in fact SPF forecasters often assign zero probability to one or more bins.[5] Hence we follow Zadora et al. (2010) and Scealy and Welsh (2011) and use a distribution which allows for values of the random vector on the boundary of the simplex. Define the sequence of indicator variables ξ_j, $j = 1, \ldots, J$ with $\xi_j = 0$ if and only if $z_j = 0$ and $\xi_j = 1$ otherwise. Define the joint distribution of $\mathbf{z} = (z_1, \ldots, z_J)$ and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_J)'$ as a zero-augmented Dirichlet distributions with probability density function

$$h(\mathbf{z}, \boldsymbol{\xi}|\boldsymbol{\theta}) = \prod_{j=1}^{J} \alpha(\nu_j(\boldsymbol{\varphi})|\varepsilon)^{\xi_j}(1 - \alpha(\nu_j(\boldsymbol{\varphi})|\varepsilon))^{1-\xi_j}\tilde{h}(\mathbf{z}|\boldsymbol{\nu}(\boldsymbol{\varphi}), \phi, \boldsymbol{\xi}). \tag{7}$$

where $\alpha(\cdot|\varepsilon)$ measures the likelihood that a forecaster will report a zero probability on the j^{th} bin, and is a function that is close to one when its argument $\nu_j(\boldsymbol{\varphi})$ is close to zero, and

[5] The same issue is present in another widely used model for compositional data that is the logistic distribution (e.g., see Aitchinson, 1982, 1986). See Pawlowsky-Glahn and Buccianti (2011), chapter 4, for a discussion on rounded, count and essential zeros in compositional data and on some alternative ways to deal with zeros.

$\tilde{h}(\mathbf{z}|\boldsymbol{\nu}(\boldsymbol{\varphi}), \phi, \boldsymbol{\xi})$ is the standard Dirichlet distribution defined on the elements of $\mathbf{z}$ that are non zero:

$$\tilde{h}(\mathbf{z}|\boldsymbol{\nu}(\boldsymbol{\varphi}), \tilde{\phi}, \boldsymbol{\xi}) = \frac{\Gamma\left(\sum_{j\in\tilde{\mathcal{J}}} \tilde{\phi}\tilde{\nu}_j(\boldsymbol{\varphi})\right)}{\prod_{j\in\tilde{\mathcal{J}}} \Gamma(\tilde{\phi}\tilde{\nu}_j(\boldsymbol{\varphi}))} \prod_{j\in\tilde{\mathcal{J}}} z_j^{\tilde{\phi}\tilde{\nu}_j(\boldsymbol{\varphi})-1}, \tag{8}$$

where $\tilde{\mathcal{J}} = \{j = 1, \ldots, J; \xi_j = 0\}$ is the set indexes of the non-zeros elements of $\mathbf{z}$, $\tilde{\phi} = \phi\kappa$ is the rescaled precision, with $\kappa = \sum_{j\in\tilde{\mathcal{J}}} \nu_j(\boldsymbol{\varphi})$, and $\tilde{\nu}_j(\boldsymbol{\varphi}) = \nu_j(\boldsymbol{\varphi})/\kappa$ for $j \in \tilde{\mathcal{J}}$ are the renormalized ν's, which take into account the fact that if a forecaster decides to report zero probability for one or more bins, she needs to adjust the probabilities associated with the other bins so that they still sum up to one.[6]

The probability $\alpha(\cdot|\varepsilon)$ of zero-valued observations is modeled as

$$\alpha(\nu|\varepsilon) = \int_0^{\varepsilon} \mathcal{B}e(x|\nu, r)dx \tag{11}$$

where $\mathcal{B}e(x|m, r)$ denotes the pdf of a Beta distribution $\mathcal{B}e(m, r)$ with mean m and precision r parameters. The parameter r is fixed at 100, and the prior for ε is such that $\alpha(\nu_j(\boldsymbol{\varphi})|\varepsilon)$ is close to one for $\nu_j(\boldsymbol{\varphi})$ less than 1%, very small for any $\nu_j(\boldsymbol{\varphi})$ larger than 5%, and virtually zero when $\nu_j(\boldsymbol{\varphi})$ is larger than 10%.[7] Figure **??** in the appendix shows the mean and the 90% coverage intervals of $\alpha(\cdot|\varepsilon)$ as a function of ν

The distribution used in the infinite mixture representation (5) is the marginal distribu-

---

[6] Note that the conditional Dirichlet satisfies some relevant properties of the unconditional Dirichlet, that are the elements of $\mathbf{z}$ and their marginal conditional means

$$\mathbb{E}(z_j|\boldsymbol{\xi}) = \frac{\tilde{\phi}\tilde{\nu}_j}{\sum_{j\in\tilde{\mathcal{J}}} \tilde{\phi}\tilde{\nu}_j(\boldsymbol{\varphi})} = \frac{\nu_j}{\sum_{j\in\tilde{\mathcal{J}}} \nu_j(\boldsymbol{\varphi})} = \tilde{\nu}_j(\boldsymbol{\varphi}),\ j \in \tilde{\mathcal{J}} \tag{9}$$

sum up to one, and their marginal conditional variances

$$\mathbb{V}(z_j|\boldsymbol{\xi}) = \frac{\tilde{\nu}_j(1-\tilde{\nu}_j)}{(\phi\sum_{j\in\tilde{\mathcal{J}}} \nu_j(\boldsymbol{\varphi}) + 1)},\ j \in \tilde{\mathcal{J}} \tag{10}$$

go to zero with $\phi \to \infty$.

[7] We chose the beta distribution because it is the marginal of a Dirichlet, but we could have chosen any other distribution satisfying the above requirements. Note that since the ξ_j are down independently from one another, in principle all ξ_j's could be equal to one, thereby violating the constraint that the sum of the z_j's is one. In practice, since the ν_j's must sum to one, the probability of such an outcome is virtually zero (that is, some ν_j is far enough from zero that $\alpha(\nu_j|\varepsilon)$ is essentially zero). Our parametrization of the beta distribution is $\mathcal{B}e(x|\nu, r) = \frac{1}{B(\nu r, (1-\nu)r)} x^{\nu r-1}(1-x)^{(1-\nu)r-1}$ with $x \in (0, 1)$, $m \in (0, 1)$ and precision $r > 0$.

tion of $\mathbf{z}$ implied by expression (8):

$$h(\mathbf{z}|\boldsymbol{\theta}) = \sum_{\boldsymbol{\xi}\in\mathcal{Z}} \prod_{j=1}^{J} \alpha(\nu_j(\boldsymbol{\varphi})|\varepsilon)^{\xi_j}(1-\alpha(\nu_j(\boldsymbol{\varphi})|\varepsilon))^{1-\xi_j} h(\mathbf{z}|\boldsymbol{\nu}(\boldsymbol{\varphi}),\phi,\boldsymbol{\xi}) \tag{12}$$

where $\mathcal{Z}$ is the set of all vectors with 0-1 binary entries ($\mathcal{Z} = \{\boldsymbol{\xi} = (\xi_1,\ldots,\xi_J)' \in \{0,1\}^J\}$), and $\boldsymbol{\theta}' = (\boldsymbol{\varphi}',\phi,\varepsilon)$. In fact, for computational reasons, we take a data augmentation approach and write the Gibbs sampler using the joint distribution $h(\mathbf{z},\boldsymbol{\xi}|\boldsymbol{\theta})$. Our infinite mixture model is then

$$h(\mathbf{z},\boldsymbol{\xi}|G) = \int h(\mathbf{z},\boldsymbol{\xi}|\boldsymbol{\theta})G(d\boldsymbol{\theta}) = \sum_{k=1}^{\infty} w_k h(\mathbf{z},\boldsymbol{\xi}|\boldsymbol{\theta}_k). \tag{13}$$

Finally, we complete the model description by specifying the base measure G_0 and the prior for the hyperparameter ψ. This is given by independent priors for μ, σ^2, ϕ, and ε that are given by a Gaussian for μ, and gamma distributions for σ^2, ϕ, and ε:

$$G_0(\boldsymbol{\theta}) = G_0(\mu,\sigma^2,\phi,\varepsilon) = \mathcal{N}(m,s^2)\mathcal{G}a(a_\sigma,b_\sigma)\mathcal{G}a(a_\phi,b_\phi)\mathcal{G}a(a_\varepsilon,b_\varepsilon). \tag{14}$$

The hyper-parameter $\psi > 0$ is driving the prior expected number of components. Large values of ψ increase the probability of introducing new components in the mixture. We assume a hierarchical prior and specify a gamma prior distribution for ψ

$$\mathcal{G}a(a_\psi,b_\psi). \tag{15}$$

We conclude this section by discussing some of the model's implications. If the forecaster never reports zero probabilities (that is, conditional on $\xi_j = 0\ \forall j$), then in expectation z_j coincides with ν_j: $\mathbb{E}[z_j|\boldsymbol{\theta}] = \nu_j(\boldsymbol{\varphi})$. Expression (13) then implies that the distribution of each z_j, conditional on $\xi_j = 0\ \forall j$, will be centered at the infinite mixture of the bin probabilities ν_j's implied by each mixture component $F(\cdot|\boldsymbol{\varphi}_k)$:

$$\mathbb{E}\left[z_j|G\right] = \sum_{k=1}^{\infty} w_k \nu_j(\boldsymbol{\varphi}_k) = \sum_{k=1}^{\infty} w_k (F(y_j|\boldsymbol{\varphi}_k) - F(y_{j-1}|\boldsymbol{\varphi}_k)). \tag{16}$$

II.C Posterior inference

Our Gibbs sampler applied to the cross section of $(\mathbf{z}_i,\boldsymbol{\xi}_i)$, $i = 1,\ldots,n$ 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 (13) can be written as the marginal distribution of

$$h(\mathbf{z}_i, \boldsymbol{\xi}_i, u_i | w_{1:\infty}, \boldsymbol{\theta}_{1:\infty}) = \sum_{k=1}^{\infty} \mathbb{I}(u_i < w_k) h(\mathbf{z}_i, \boldsymbol{\xi}_i | \boldsymbol{\theta}_k) \tag{17}$$

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 $\mathbf{z}_i$ and $\boldsymbol{\xi}_i$ given u_i, the weights and the atoms, is

$$h(\mathbf{z}_i, \boldsymbol{\xi}_i | u_i, w_{1:\infty}, \boldsymbol{\theta}_{1:\infty}) = \frac{1}{h(u_i | w_{1:\infty})} \sum_{k \in A(u_i | w_{1:\infty})} h(\mathbf{z}_i, \boldsymbol{\xi}_i | \boldsymbol{\theta}_k), \tag{18}$$

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 (13), expression (18) 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

$$L(\mathbf{z}_{1:n}, \boldsymbol{\xi}_{1:n} | u_{1:n}, d_{1:n}, v_{1:\infty}, \boldsymbol{\theta}_{1:\infty}) = \prod_{i=1}^{n} \mathbb{I}_{\{u_i < w_{d_i}\}} h(\mathbf{z}_i, \boldsymbol{\xi}_i | \boldsymbol{\theta}_{d_i}) \tag{19}$$

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 (4). The complete Gibbs sampler is given in Appendix A.[8]

II.D Posterior consistency

Let $\mathcal{X}$ be the sample space with elements x and Θ the space of the mixing parameter θ and Υ the space of the hyperparameter v. Let $\mathcal{F}(\mathcal{X})$ be the space of probability densities on $\mathcal{X}$, $\mathcal{M}(\Theta)$ the space of probability measures on the mixing probability space Θ, and G the mixing distribution on Θ with density g and a prior Π on $\mathcal{M}(\Theta)$. Denote a prior for v

[8] See also Escobar (1994) and Ishwaran and James (2001) for a Gibbs sampler based on the Polya-urn representation of the Dirichlet process and Ishwaran and Zarepour (2000) for a Gibbs sampler based on a truncation of the infinite number of mixture components. Finally, Papaspiliopoulos and Roberts (2008) proposed an exact simulation algorithm based on retrospective sampling.

by π and the weak support of π by $\text{supp}(\pi)$ with π independent on G. Let $K(x;\theta,\upsilon)$ be a probability kernel on $\mathcal{X} \times \Theta \times \Upsilon$, then a type II mixture prior (see Wu and Ghosal (2009a)) Π^* on $\mathcal{F}(\mathcal{X})$ induced by Π, π and $K(x;\theta,\upsilon)$ is defined via the map

$$(\upsilon, G) \mapsto f_{\upsilon,G}(x) = \int K(x;\theta,\upsilon) dG(\theta) \tag{20}$$

In our model the kernel is

$$K(x;\theta,\upsilon) = h(\mathbf{z}, \boldsymbol{\xi}|\boldsymbol{\theta}) \tag{21}$$

where $\theta = (\mu, \sigma^2)$, $\upsilon = (\varepsilon, \phi)$, $x = (\mathbf{z}, \boldsymbol{\xi})$. The sample space is $\mathcal{X} = \{0,1\}^J \times \Delta^J$ where Δ^J denote the J-dimensional simplex. The random mixing distribution Π is given by the Dirichlet process prior $(\mu,\sigma)|G \sim G$ where $G \sim \mathcal{DP}(\psi, G_0)$ with $G_0 = \mathcal{N}(m, s^2)\mathcal{G}a(a_\sigma, b_\sigma)$. The hyperprior π is $\mathcal{G}a(a_\varepsilon, b_\varepsilon)\mathcal{G}a(a_\phi, b_\phi)$.

Let us denote with $KL(f,g)$ the Kullback-Leibler (KL) divergence between $f, g \in \mathcal{F}(\mathcal{X})$, i.e. $KL(f,g) = \int f \log(f/g)$. Then, Schwartz theorem states that weak posterior consistency at a "true density" $f_0 \in \mathcal{F}(\mathcal{X})$ holds if the prior assigns positive probabilities to the Kullback-Leibler neighborhoods of f_0, where a Kullback-Leibler neighborhood of size η of a density $f_0 \in \mathcal{F}(\mathcal{X})$ is defined as $\mathcal{K}_\eta(f_0) = \{f \in \mathcal{F}(\mathcal{X}) | KL(f_0, f) < \eta\}$. By convention, we say that the KL property holds at $f_0 \in \mathcal{F}(\mathcal{X})$ or f_0 is in the KL support of Π^*, and write $f_0 \in KL(\Pi^*)$ if $\Pi^*(\mathcal{K}_\eta(f_0)) > 0$ for every $\eta > 0$.

We shall notice that in our model the kernel $h(\mathbf{z}, \boldsymbol{\xi}|\boldsymbol{\theta}) = \tilde{h}(\mathbf{z}|\boldsymbol{\theta}, \boldsymbol{\xi})p(\boldsymbol{\xi}|\alpha(\boldsymbol{\theta}))$ decomposes as the product of continuous part $\tilde{h}(\mathbf{z}|\boldsymbol{\theta}, \boldsymbol{\xi})$ and discrete part $p(\boldsymbol{\xi}|\alpha(\boldsymbol{\theta}))$, where α is the probability $\xi_j = 1$ and $1-\alpha$ the complementary probability associated with $\xi_j = 0$. From Eq. 11, one can see that the kernel is a rounded kernel with probability mass written as integrals of density functions, that is $\alpha(f) = \int_0^\varepsilon f(x|\boldsymbol{\theta})dx$ and $(1-\alpha(f)) = \int_\varepsilon^1 f(x|\boldsymbol{\theta})dx$, with $f(x|\boldsymbol{\theta})$ the density of a beta distribution $\mathcal{B}e(\nu(\boldsymbol{\theta}), r)$. Thus, the probability mass writes as a mapping $g(f) = (\alpha, 1-\alpha)$ from the set $\mathcal{F}([0,1])$ of densities on $[0,1]$ to the set $\mathcal{F}(\{0,1\})$ of probability masses on $\{0,1\}$. The mapping $g : \mathcal{F}([0,1]) \mapsto \mathcal{F}(\{0,1\})$ maintains the KL neighborhoods of $\mathcal{F}([0,1])$. In the following lemma, for the shake of simplicity and without loss of generality we state the result for the univariate case.

Lemma 1. Assume $p_0 = (\alpha_0, 1-\alpha_0) \in \mathcal{F}(\{0,1\})$ and $f_0 \in \mathcal{F}([0,1])$ such that $p_0 = g(f_0)$. Let $\mathcal{K}_\eta(f_0)$ a KL neighborhood of size η around f_0. Then the image $g(\mathcal{K}_\eta(f_0))$ contains values $p \in \mathcal{F}(\{0,1\})$ in the KL neighborhood $\mathcal{K}_\eta(p_0)$ of p_0.

Proof. The proof follows the same lines of the proof of Lemma 1 in Canale and Dunson (2011). We have

$$
\begin{aligned}
KL(f_0, f) &= \int_0^1 f_0(x)\log(\frac{f_0(x)}{f(x)})dx = \int_0^{\varepsilon} f_0(x)\log(\frac{f_0(x)}{f(x)})dx + \int_{\varepsilon}^1 f_0(x)\log(\frac{f_0(x)}{f(x)})dx \\
&\geq \int_0^{\varepsilon} f_0(x)dx\left(\log(\int_0^{\varepsilon} f_0(x)dx) - \log(\int_0^{\varepsilon} f(x)dx)\right) + \\
&\quad \int_{\varepsilon}^1 f_0(x)dx\left(\log(\int_{\varepsilon}^1 f_0(x)dx) - \log(\int_{\varepsilon}^1 f(x)dx)\right) \\
&= \alpha_0\log(\frac{\alpha_0}{\alpha}) + (1-\alpha_0)\log(\frac{1-\alpha_0}{1-\alpha})
\end{aligned}
$$

Thus if $f \in \mathcal{K}_\eta(f_0)$, $\eta > KL(f_0, f) \geq KL(p_0, p)$ then $p \in \mathcal{K}_\eta(p_0)$.

Since f is parametrized in $\boldsymbol{\theta}$, the previous lemma allows us to show the weak posterior consistency looking at the probability assigned to the neighborhoods of f_0 by the prior distributions defined on Θ. Since for in our model the conditions in Theorem 1 and Lemma 3 of Wu and Ghosal (2009b,a) are satisfied then $\Pi^*(\mathcal{K}_\eta(f_0)) > 0$.

II.E Prior Specification for the Empirical Application

TBW

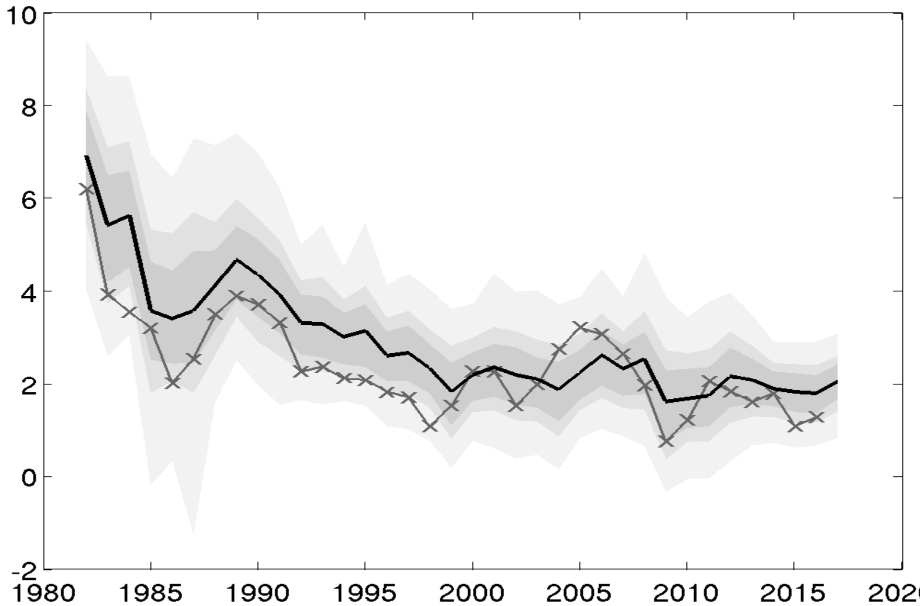
III Results

III.A The Conquest of Inflation Credibility in the U.S.

- Figure 1 shows the aggregate subjective distribution of 2-year ahead inflation resulting from our estimation. Specifically, if $F_{t,i}(y) = \sum_{k=1}^{\infty} w_{ik}^{(t)} F(y|\theta_k^{(t)})$ is the estimated underlying subjective distribution for forecaster i (recall that the $w_{ik}^{(t)}$ are the weights of the infinite mixture associated with forecaster i in period t) we show the mean, 68, and 95 percent posterior coverage intervals (gray line with markers is actual inflation) associated with

$$\bar{F}_t(y) = \frac{1}{n}\sum_{i=1}^{n} F_{t,i}(y)$$

Figure 1: The Aggregate Subjective Distribution of 2-Year Ahead Inflation



Note: The shaded areas show the 68 and 90 percent posterior coverage intervals.

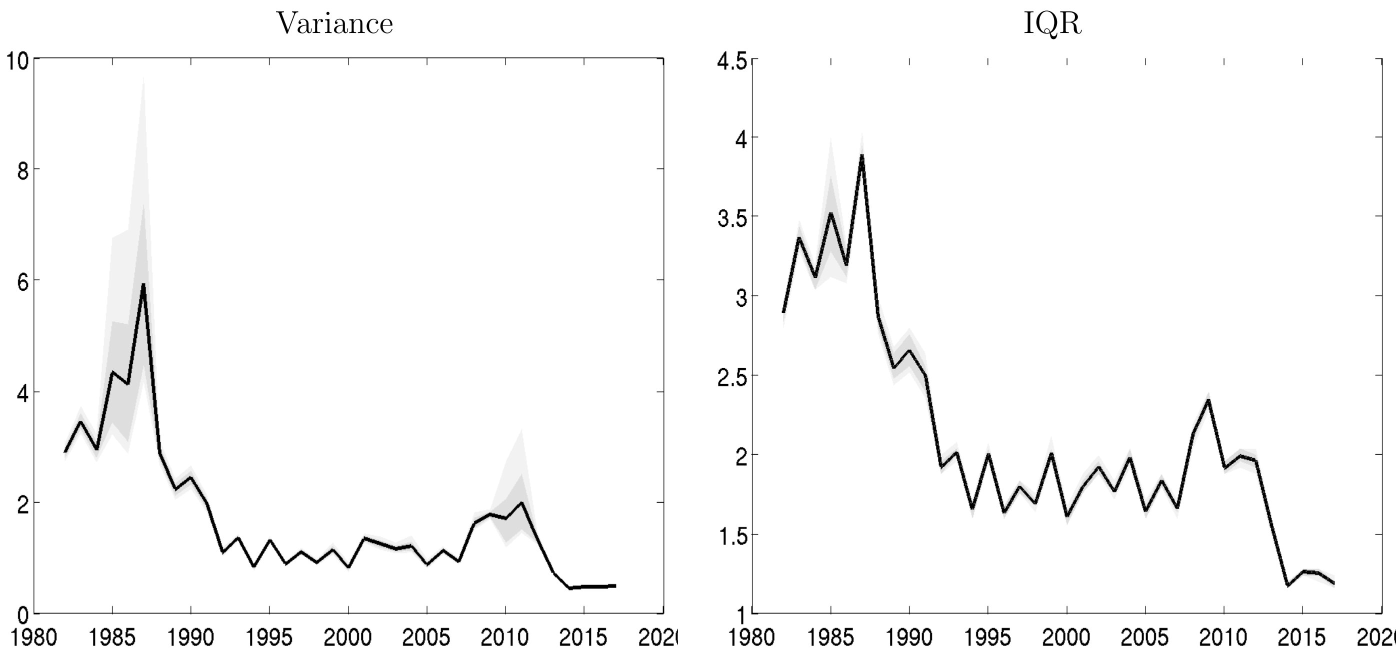
- Tom Sargent's famous work on "The Conquest of American Inflation" is about first moments — the decline over time in both realized inflation and mean inflation expectations. This paper is about the second moment of the forecast distribution, the variance, which reflects the forecasters' belief of the extent to which the central bank can keep inflation under control.

- The variance of the distribution narrows substantially over time. This narrowing is what we refer to as the "conquest of inflation credibility" in the U.S. This is evident in Figure 2 below. Both drop dramatically in the early nineties, and drop again after the announcement of the Federal Reserve's long term inflation goal in early 2012. The rest of the paper documents the drivers of this change in the second moment of the subjective distribution. We find that

  1. The fall in disagreement plays a truly minor role. It is relevant only in the eighties and even then it is due to one or two outliers.
  2. Almost all of the decline in the variance of the aggregate distribution is due to a decline in average uncertainty. This is a result robust to omitting outliers.
  3. The decline in uncertainty is not at all homogeneous across forecasters. In fact, we show that for about half of the forecasters inflation uncertainty was very low to start with, even in the eighties. The "conquest of inflation credibility" therefore

amounts to "high-uncertainty" forecasters becoming convinced that inflation was under control.

Figure 2: Variance and Interquartile Range of the Aggregate Subjective Distribution of 2-Year Ahead Inflation



Note: The shaded areas show the 68 and 90 percent posterior coverage intervals.

- Figure 3 uses the law of total variance to decompose the variance of the aggregate distribution into uncertainty (dashed line – average variance across forecasters) and disagreement (dotted line – variance of mean forecasts). Disagreement seems to play a large role in the mid-eighties, and a fairly minor role relative to uncertainty otherwise.

- Figure 4 shows disagreement and uncertainty, together with the measures computed throwing away just two outliers per cross-section (lines with markers). For disagreement we throw away the forecasters with highest and lowest mean, and for uncertainty the forecasters with highest and lowest variance (removing the two highest variance forecasters yields very similar results). Clearly, much of disagreement in the 80s is spurious, as it it due to just two individuals. The decline in uncertainty is robust. Figure 5 shows the variance of the aggregate distribution after throwing away the two outliers (highest/lowest mean forecasters): uncertainty is clearly the key driver.

- Figure 6 provides more evidence of the reduction in uncertainty over the sample period, especially for "high-uncertainty" forecasters. It shows the predictive distributions

Figure 3: Variance Decomposed into Uncertainty and Disagreement

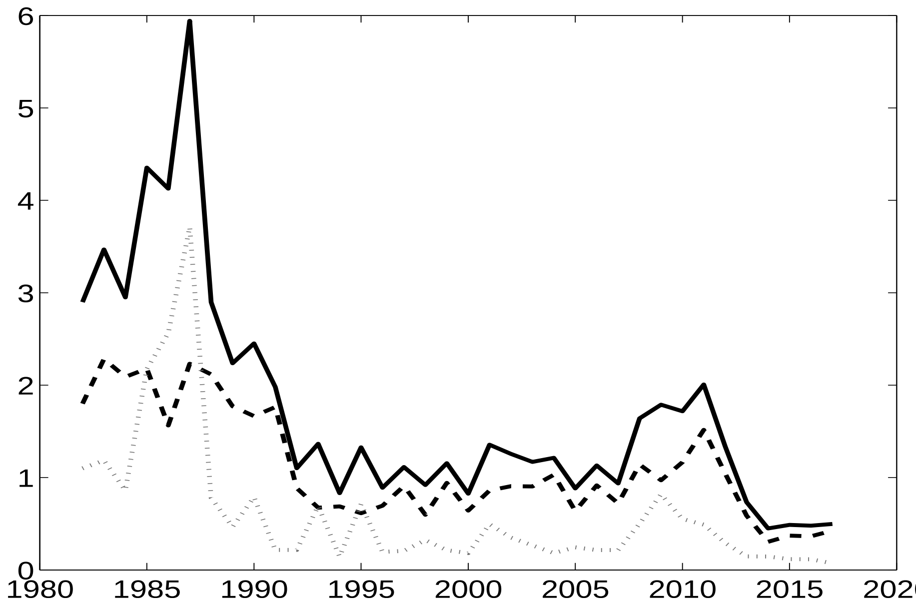
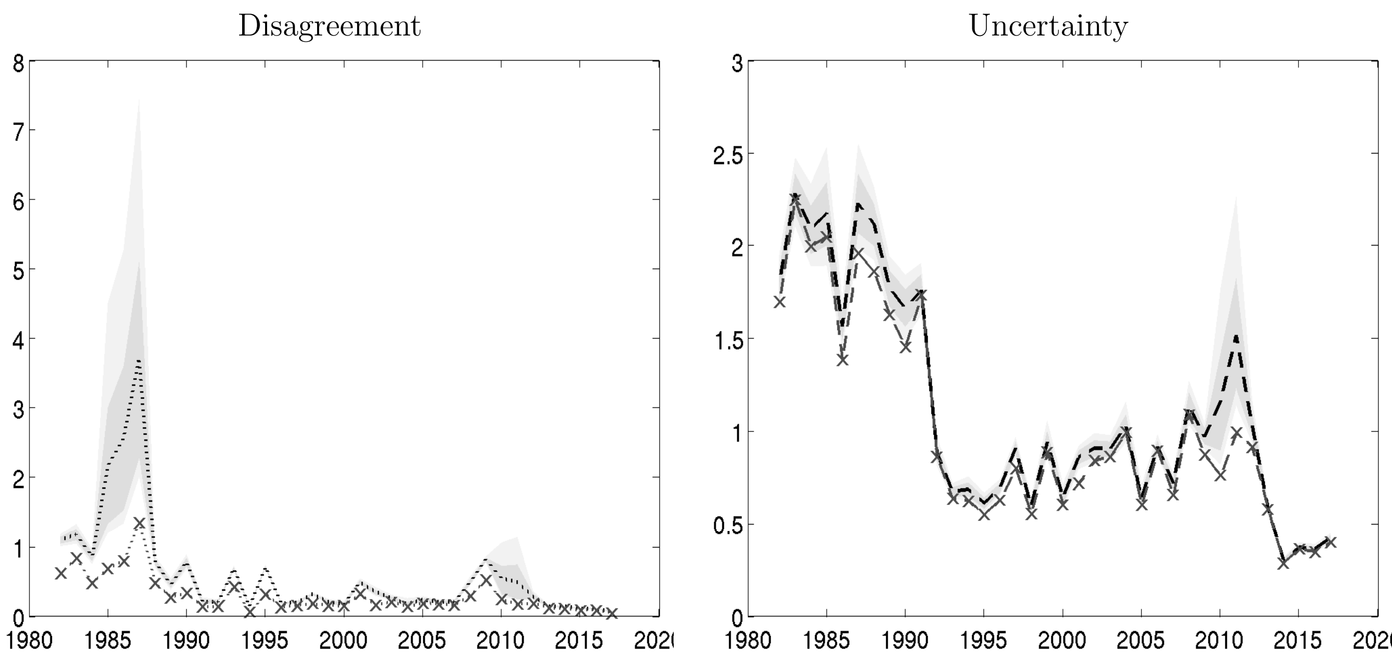


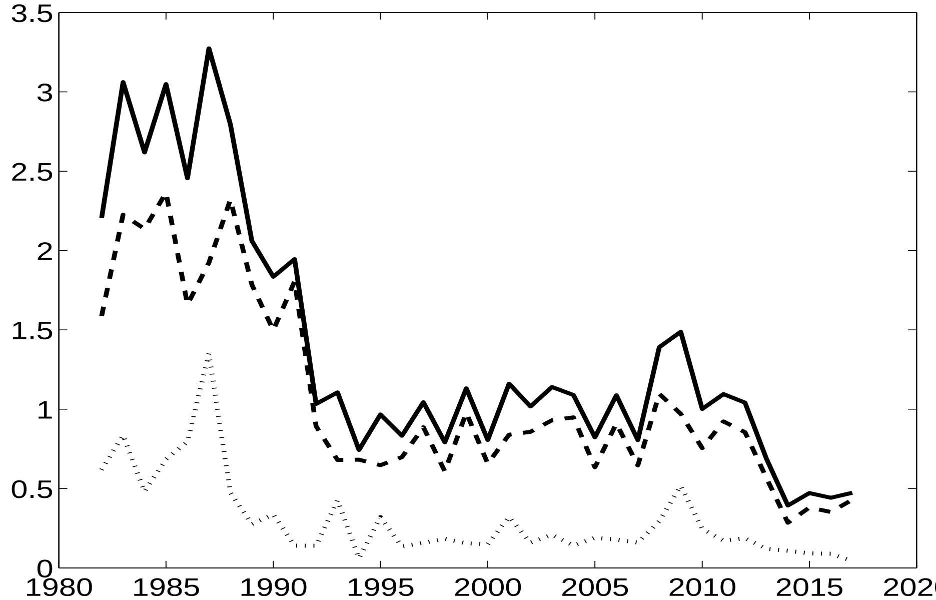
Figure 4: Uncertainty and Disagreement: With and Without Outliers



Note: The shaded areas show the 68 and 90 percent posterior coverage intervals.

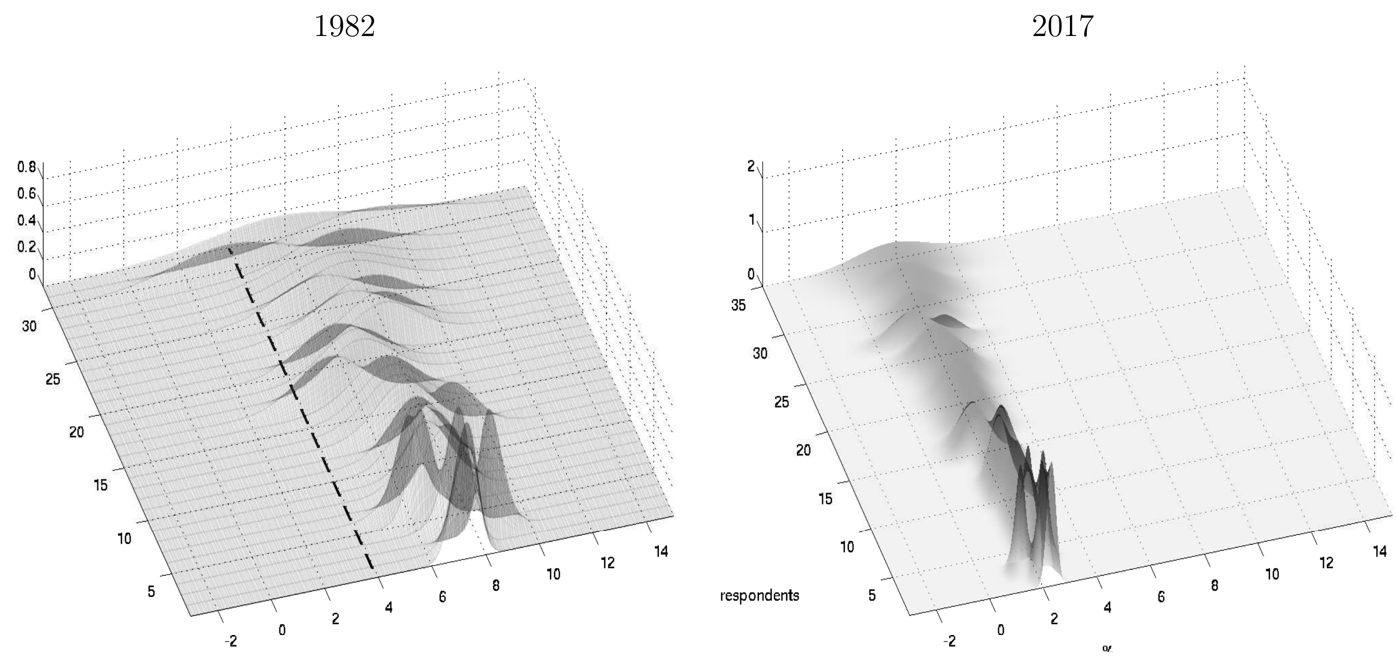
across forecasters, ranked by variance, at the beginning (1982) and at the end of our sample (2017). All subjective distribution of course shift to the left – on average predictive inflation is lower — but the most striking difference between the two plots is the

Figure 5: Variance Decomposed into Uncertainty and Disagreement – Without Outliers



reduction in uncertainty for all forecasters, but especially for high variance forecasters.

Figure 6: Predictive Distributions Across Forecasters: 1982 vs 2017

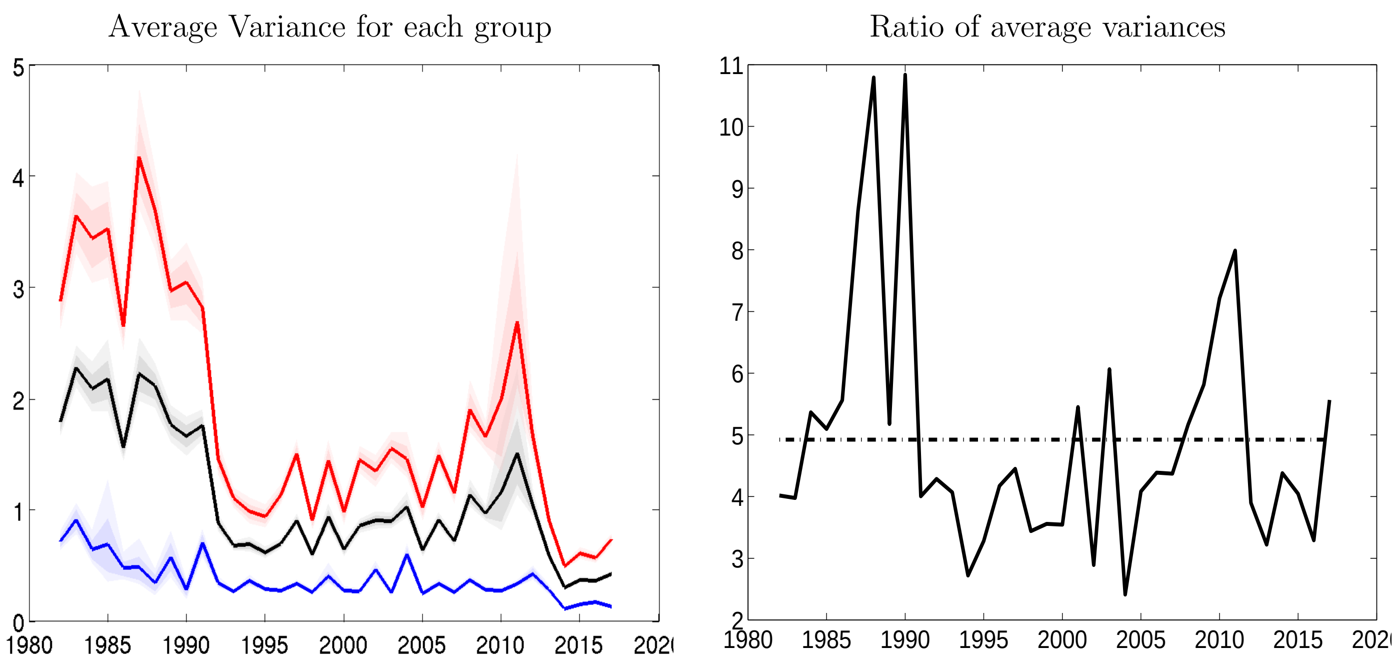


III.B Heterogeneity in Uncertainty

- Figure 7 splits for each cross sections the forecasters in two — high and low variance. The left panel and shows the evolution of average variance for each group. The right panel shows the evolution of the ratio of the two.

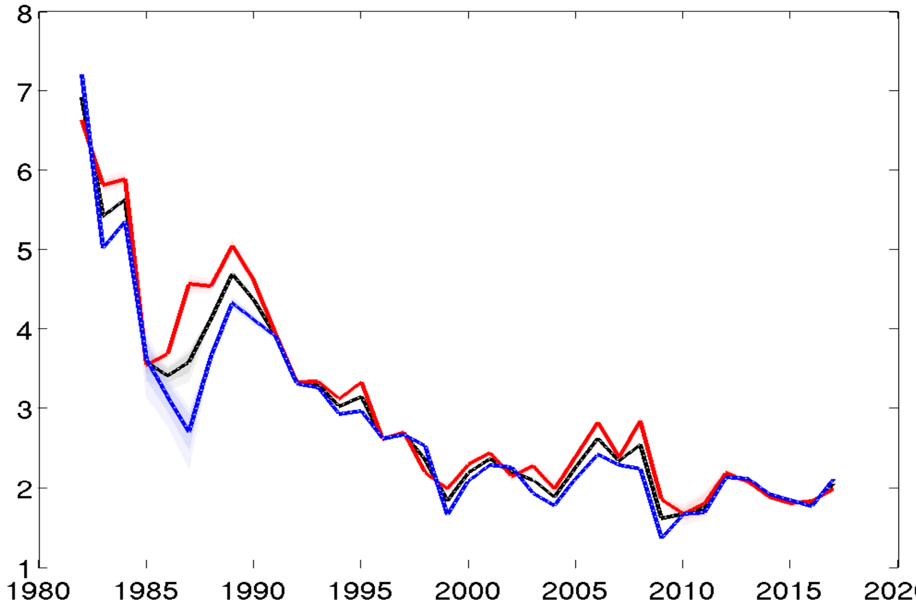
  1. Low variance forecasters have pretty low variance throughout the sample. They were convinced from the beginning that inflation was "under control". High variance forecasters were definitely not convinced in the eighties, but their variance drops in the early nineties. This drop drives the fall in the variance of the aggregate forecast distribution – the "conquest of inflation credibility".
  2. Heterogeneity in the variance is large: At times high variance forecasters's uncertainty is almost ten times as large as that of that for low variance forecasters (right panel)

Figure 7: Heterogeneity in Uncertainty



- Figure 8 asks whether high and low variance forecasters also have different mean projections. Specifically, it shows the mean forecasts, averaged across high (red) and low (blue) variance forecasters. It makes the point that high and low variance forecasters do not have different mean projections on average. High variance forecasters tend to have slightly higher mean projections, but the difference is generally not large, except perhaps for the mid-eighties.

Figure 8: Differences in Mean Forecasts: High vs Low Uncertainty Forecasters

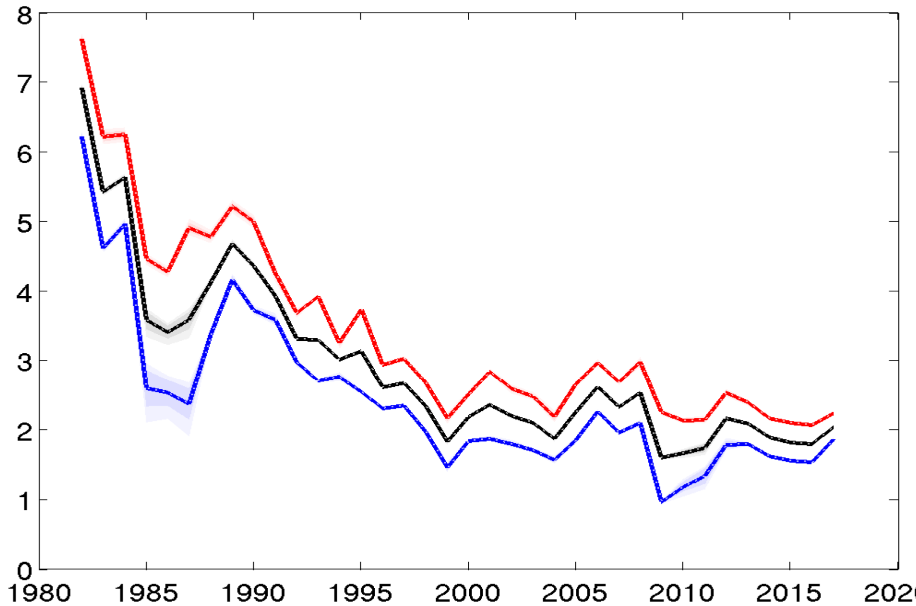


- Figure 9 investigates heterogeneity in the mean projections, and shows that it is arguably much less interesting than heterogeneity in uncertainty. Specifically, for each cross sections we rank forecasters according to their mean inflation forecasts and split forecasters into two groups — high and low mean. Figure 9 shows the average mean projections for each group. It shows that by and large the average mean projections for high (red) and low (blue) mean forecasters move in parallel as they decline over time. This is another way of making the point that changes in disagreement have not been a key feature of the evolution of the distribution of subjective forecasts for the SPF.

- Figure 10 describes the evolution over time in forecasters' heterogeneity. It shows the kernel estimates of cross-sectional distribution of variances (left panel) and means (right panel) from 1982 to 2017.

- The left panel of Figure 10 makes two points. First, throughout the sample there is always a substantial mass of forecasters with low variance. Second, the main change over time in the cross sectional distribution concerns the right tail. This is very heavy

Figure 9: Differences in Mean Forecasts: High vs Low Mean Forecasters



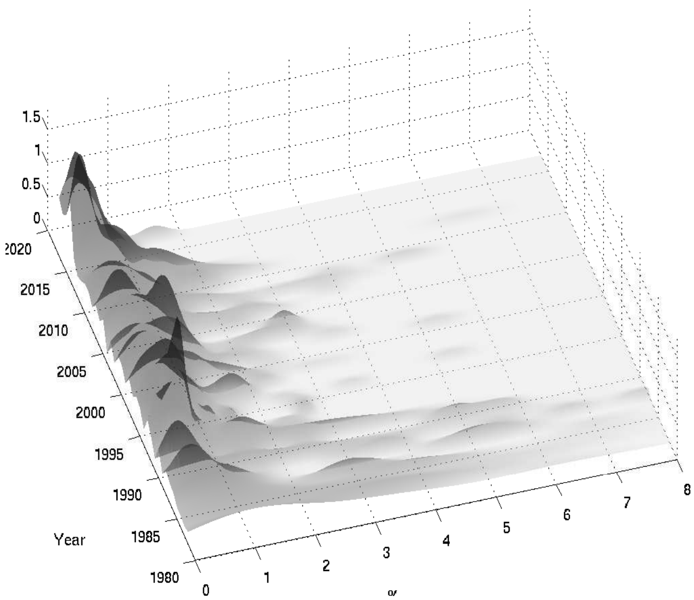
in in the eighties and early nineties, but becomes less and less important over time, and virtually disappears toward the end of the sample.

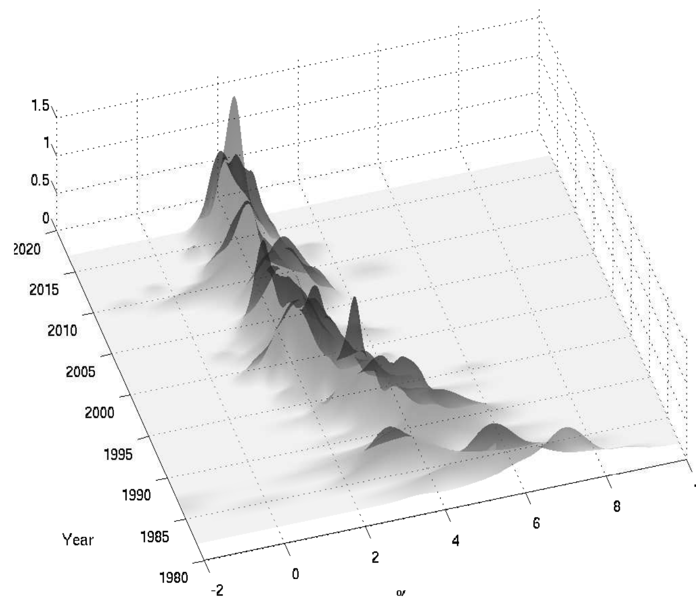
- In sum, the cross-sectional distribution of variances changed quite a bit from the beginning to the end of the sample, consistently with what we have shown so far. The cross-sectional distributions of the means (right panel) by and large is very similar across the entire sample, however, except of course for the decrease in the mean (and for some erratic behaviour in the mid-eighties).

III.C Back to the Seventies?

- Next, we study the probability of a return to the inflation levels of the seventies, as viewed from the perspective of the SPF forecasters. We think this is an interesting question because it provides indirect evidence on the empirical appeal of recent DSGE models with regime switching (e.g. Bianchi and Ilut (2015), Bianchi and Melosi (2014); see Sims and Zha (2006) for a VAR version). In these models one of the regime can indeed be characterized as a "back to the seventies" regime. Do forecasters truly entertain such a possibility?

Figure 10: Evolution Over Time in the Cross-Sectional Distribution of Variances and Means





- Figure 11 shows the probability that inflation is at least as high as 6.9% , the average year-over-year inflation rate in the seventies.

- The first row of the panel shows these probabilities computed using the aggregate forecast distribution $1-\bar{F}_t(y)$. Not surprisingly, these probabilities were elevated (more than 50%) in the early eighties, but fell substantially thereafter becoming virtually negligible (left panel). The right plot zooms on the period since 2000, and shows that $1-\bar{F}_t(6.9)$ is always significantly less than 0.5%.

- In the other rows we rank $1-F_{i,t}(6.9)$ and show the values for the top 25th, 10th, and 5th quintile of the cross-sectional distribution. These probabilities are all above $1-\bar{F}_t(6.9)$ in the eighties, indicating that a substantial fraction of forecasters were convinced inflation was going to return to high levels. Interestingly, this is not the case in the recent period. For instance, $1-F_{top\ 25,t}(6.9) < 1-\bar{F}_t(6.9)$ for $t \geq 2000$, suggesting that at least 75% of forcasters believe that a return to the seventies is highly unlikely. One has to get to the 10th quantile of the cross-sectional distribution of $1-F_{i,t}(6.9)$ to see probabilities around 1%, and to the 5th quantile (basically, the second highest forecaster, given the size of the cross-section) to find any substantial probability.

- In sum, we find that only a fringe of forecasters place any significant probability of the possibility of a return to the seventies.

III.D Deflation Fear

- The Great Recession and its aftermath, with very large output gaps and conventional monetary policy constrained by the zero lower bound, rose the specter of deflation, which the U.S. economy experienced in the 1930s following the Great Depression. Again, we investigate the likelihood of deflation from the perspective of the professional forecasters.

- The first row of the panel in Figure 12 shows the probability of deflation (that is, $P\{y \leq 0\}$) computed using the aggregate forecast distribution $\bar{F}_t(y)$. Interestingly, this probability was non negligible in the mid-eighties, which is further evidence of the high uncertainty during this period. In the recent period the likelihood of deflation rises to almost ten percent in the aftermath of the Great Recession but falls thereafter to about 1 percent.

- In the other rows we rank $F_{i,t}(0)$ and show the values for the top 25th, 10th, and 5th quintile of the cross-sectional distribution. The likelihood of deflation for the 25th quartile of the cross sectional distribution is not very different from that of the average distribution, indicating that the distribution of $F_{i,t}(0)$ is quite right-tailed.

- For forecasters in the 5th and 10th quantile the likelihood of deflation becomes large in 2009—almost 50%—but decreases rapidly thereafter. Currently this probability is larger than that associated with a return to the seventies, which is easily explained by the fact that inflation is currently much closer to zero than to 6 percent, but is not particularly large.

Figure 11: Evolution Over Time of High Inflation Probabilities

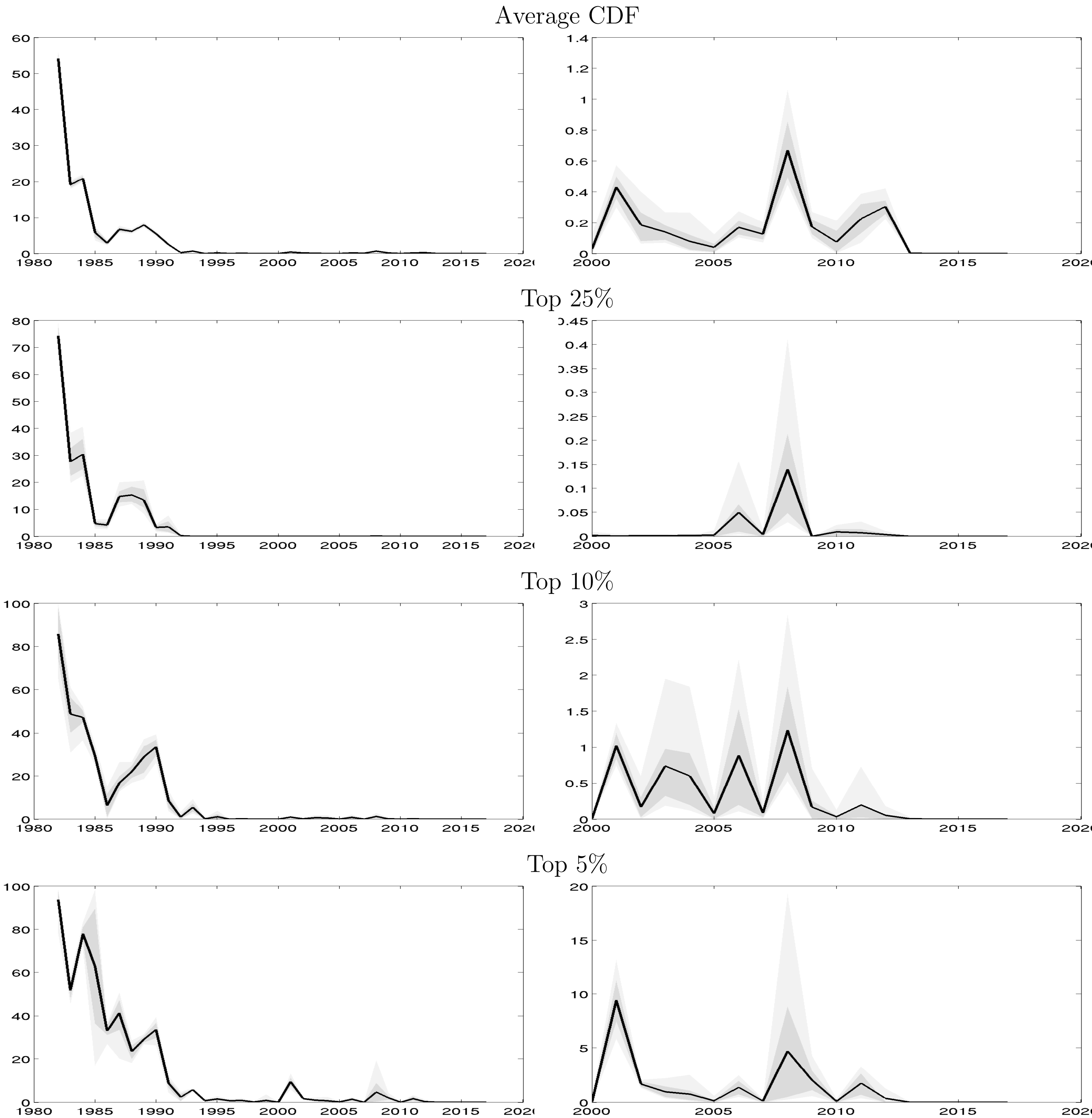
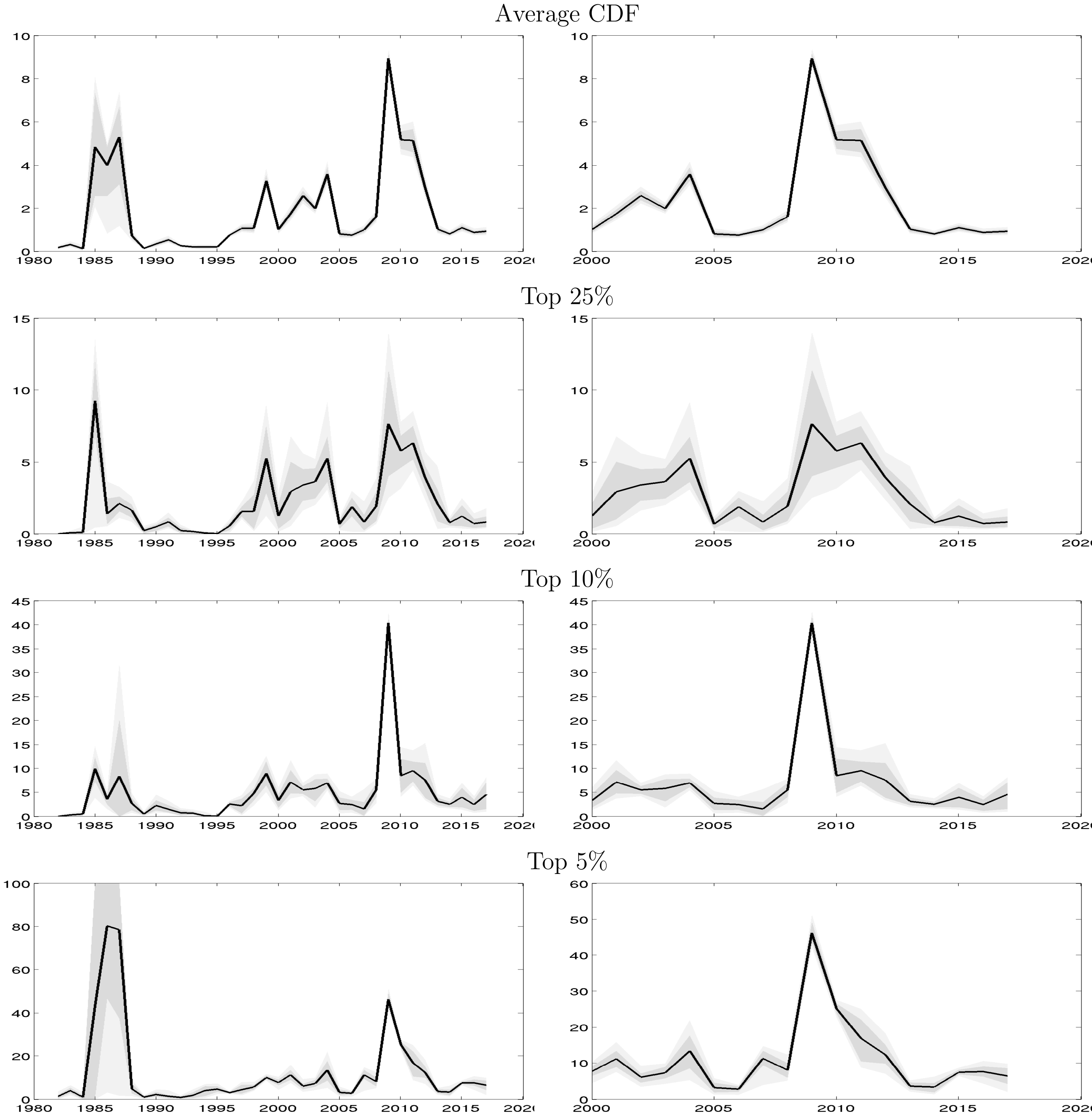


Figure 12: Evolution Over Time of Deflation Probabilities



IV Conlusions

TBW

References

Aitchinson, J., "The statistical analysis of compositional data," *Journal of the Royal Statistical Society Series, Series B*, 1982, *44*, 139–177.

_ , *The Statistical Analysis of Compositional Data*, London: Chapman & Hall, 1986.

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

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

Bianchi, Francesco and Cosmin Ilut, "Monetary/fiscal policy mix and agents' beliefs," *National Bureau of Economic Research*, 2015.

_ **and Leonardo Melosi**, "Escaping the Great Recession," Technical Report, National Bureau of Economic Research 2014.

Blei, David M., Thomas L. Griffiths, and Michael I. Jordan, "The nested chinese restaurant process and bayesian nonparametric inference of topic hierarchies," *Journal of the ACM (JACM)*, 2010, *57* (2), 7.

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.

Canale, Antonio and David B. Dunson, "Bayesian kernel mixtures for counts," *Journal of the American Statistical Association*, 2011, *106* (496), 1528–1539.

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

Dominitz, Jeff and Charles F Manski, "Using expectations data to study subjective income expectations," *Journal of the American statistical Association*, 1997, *92* (439), 855–867.

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.

Escobar, M. D., "Estimating normal means with a Dirichlet process prior," *Journal of the American Statistical Association*, 1994, *89*, 268–277.

_ **and M. West**, "Bayesian density estimation and inference using mixtures," *Journal of the American Statistical Association*, 1995, *90*, 577–588.

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

Ghosh, Jayanta K and RV Ramamoorthi, *Bayesian nonparametrics*, Vol. 16, Springer Series in Statistics. Springer-Verlag, New York, 2003.

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.

_ **and** _ , "Stick-breaking autoregressive processes," *Journal of econometrics*, 2011, *162* (2), 383–396.

_ **, 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.

Griffiths, Thomas L. and Joshua B. Tenenbaum, "Optimal predictions in everyday cognition," *Psychological science*, 2006, *17* (9), 767–773.

Hannah, Lauren A, David M Blei, and Warren B Powell, "Dirichlet process mixtures of generalized linear models," *Journal of Machine Learning Research*, 2011, *12* (Jun), 1923–1953.

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

Hjort, Nils Lid, Chris Holmes, Peter Müller, and Stephen G Walker, *Bayesian nonparametrics*, Vol. 28, Cambridge University Press, 2010.

Ishwaran, H. and L. F. James, "Gibbs sampling methods for stick-breaking priors," *Journal of the American Statistical Association*, 2001, *96*, 161–173.

_ **and M. Zarepour**, "Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models," *Biometrika*, 2000, *87*, 371–390.

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

Kalli, M., J. E. Griffin, and S. G. Walker, "Slice sampling mixture models," *Statistics and Computing*, 2011, *21*, 93–105.

Koop, Gary, "Parametric and nonparametric inference in equilibrium job search models," *Advances in Econometrics*, 2003, *23*, 217–244.

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

Mitra, Riten and Peter Müller, *Nonparametric Bayesian Inference in Biostatistics*, Springer, 2015.

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

Papaspiliopoulos, O. and G. Roberts, "Retrospective Markov chain Monte Carlo for Dirichlet process hierarchical models," *Biometrika*, 2008, *95*, 169–186.

Pawlowsky-Glahn, V. and A. Buccianti, *Compositional Data Analysis: Theory and Applications*, Wiley, 2011.

_ **, J. J. Egozcue, and R. Tolosana-Delgado**, *Modeling and Analysis of Compositional Data*, Wiley, 2015.

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.

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

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

Sims, Christopher A. and Tao Zha, "Were There Regime Switches in U.S. Monetary Policy?," *American Economeic Review*, 2006, *96* (1), 54 – 81.

Walker, S. 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" [MR2399197]," *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.

A The Gibbs Sampler

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

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

   Now note that since $v_{\bar{d}+1:\infty}$ do not enter the likelihood (19) – that is, the term within the first parenthesis – they can be easily integrated out resulting in

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

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

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

   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) $u_{1:n}|v_{1:\bar{d}}, d_{1:n}, \boldsymbol{\theta}_{1:\infty}, \psi, \mathbf{z}_{1:n}, \boldsymbol{\xi}_{1:n}$.

The likelihood (19), seen as a function of each u_i, $i = 1, \ldots, n$, is simply a uniform distribution over $[0, w_{d_i}]$. Hence

$$\pi(u_i|\ldots) \propto \frac{1}{w_{d_i}} \mathbb{I}(u_i < w_{d_i}). \tag{A-2}$$

(c) $v_{\bar{d}+1:\infty}|u_{1:n}, v_{1:\bar{d}}, d_{1:n}, \boldsymbol{\theta}_{1:\infty}, \psi, \mathbf{z}_{1:n}, \boldsymbol{\xi}_{1:n}$.

Again, $v_{\bar{d}+1:\infty}$ do not enter the likelihood (19), so samples from those v_k with $k > \bar{d}$ are simply obtained by drawing from the prior $Beta(1, \psi)$:

$$\pi(v_k|u_{1:n}, d_{1:n}, \ldots) \propto (1 - v_k)^{\psi-1}. \tag{A-3}$$

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 (19) reveals that those mixtures for which $w_k < u_i$ will never be used, at least given the the draw for u_i. Let $\bar{n}_i$ the smallest integer such that $\sum_{k=1}^{\bar{n}_i} w_k > 1 - u_i$. 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 $\bar{n} = \max\{\bar{n}_i, i = 1, \ldots, n\}$. 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. $\boldsymbol{\theta}_{1:\infty}|v_{1:\infty}, u_{1:n}, d_{1:n}, \psi, \mathbf{z}_{1:n}, \boldsymbol{\xi}_{1:n}$

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

$$\pi(\boldsymbol{\theta}_k|\ldots) \propto \left(\prod_{i \in \mathcal{D}_k} h(\mathbf{z}_i, \boldsymbol{\xi}_i|\boldsymbol{\theta}_k) \right) G_0(\boldsymbol{\theta}_k) \tag{A-4}$$

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

$$\boldsymbol{\theta}^* \sim \mathcal{N}(\boldsymbol{\theta}^{(j-1)}, \Upsilon^{(j)}) \tag{A-5}$$

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

$$\xi^{(j)} = \xi^{(j-1)} + \gamma^{(j)}(\hat{\alpha}^{(j-1)} - \bar{\alpha}) \tag{A-6}$$

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 draws from the base measure (14).

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

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

$$\pi(d_i | \ldots) \propto \mathbb{I}(u_i < w_{d_i}) h(\mathbf{z}_i, \boldsymbol{\xi}_i | \boldsymbol{\theta}_{d_i}) \tag{A-7}$$

with $d_i \in \{1, \ldots, \bar{n}_i\}$. 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 $h(\mathbf{z}_i, \boldsymbol{\xi}_i | \boldsymbol{\theta}_k)$.

4. $\psi | v_{1:\infty}, u_{1:n}, d_{1:n}, \boldsymbol{\theta}_{1:\infty}, \mathbf{z}_{1:n}, \boldsymbol{\xi}_{1:n}$???

B Data description

We focus on the Survey of Professional Forecasters, managed since 1990 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, who include university professors and private-sector macroeconomic researchers, are asked to predict American GDP, inflation, unemployment, interest rates, and other macroeconomic variables. 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. The composition of the panel changes gradually over time, with individual members providing forecasts for about six years on average. The lowest number of respondents across vintages was 14 and the highest was 48.

We restrict our attention to the variable year-over-year GDP deflator inflation (PRPGDP) over the following year for the sample 1982Q1-2015Q1. Our sample choice is explained by the following reasons. The survey began asking forecasters for their annual (rather than quarterly) predictions in the third quarter of 1981. Probabilistic forecasts are only for annual changes. Moreover, the horizon varies across quarters. Indeed, the survey asks inflation growth for the current and next year, implying the longest horizon is eight quarters (two-years) ahead in the first quarter; seven quarters ahead in the second quarter; six quarters ahead in the third quarter; five quarters ahead in the fourth quarters. We restrict our study to the 2-years ahead horizons and collect predictions in the first quarter of each year for a total of 34 predictions.

The intervals in which respondents place probabilities have changed over the years. From the third quarter of 1981 through the end of 1991, there were 6 intervals, and after 1991, there were 10 intervals, see Table A-1.

Figure **??** show the density forecast over time of the aggregate survey taking an average across individual probabilistic predictions.

Table A-1: Probability Ranges

	Ranges (Year-over-Year Percent Changes, Percentage Points)			
Bins	1981:Q3 to 1985:Q1	1985:Q2 to 1991:Q4	1992:Q1 to 2013:Q4	2014:Q1 to Present
1	> 12	> 10	> 8	> 4
2	10 to 11.9	8 to 9.9	7 to 7.9	3.5 to 3.9
3	8 to 9.9	6 to 7.9	6 to 6.9	3 to 3.4
4	6 to 7.9	4 to 5.9	5 to 5.9	2.5 to 2.9
5	4 to 5.9	2 to 3.9	4 to 4.9	2 to 2.4
6	< 4	< 2	3 to 3.9	1.5 to 1.9
7			2 to 3.9	1 to 1.4
8			1 to 1.9	0.5 to 0.9
9			0 to 0.9	0 to 0.4
10			< 0	< 0

Note: