A Bayesian Approach for Inference on Probabilistic Surveys

Marco Del Negro
Federal Reserve Bank of New York
Roberto Casarin
Universitá Ca' Foscari of Venice
Federico Bassetti
Politecnico, Milan

NBP Workshop on Forecasting, Warsaw, November 2018

Disclaimer: **The views expressed are ours and do not necessarily reflect those of the Federal Reserve Bank of New York or the Federal Reserve System**

Goals of the Paper

1. Methodological contribution: Inference on the **forecast distribution** underlying probabilistic surveys using a Bayesian non-parametric approach

2. Application to U.S. Survey of Professional Forecasters' projections of inflation (and output growth): Has the variance of inflation narrowed—from the perspective of the SPF?

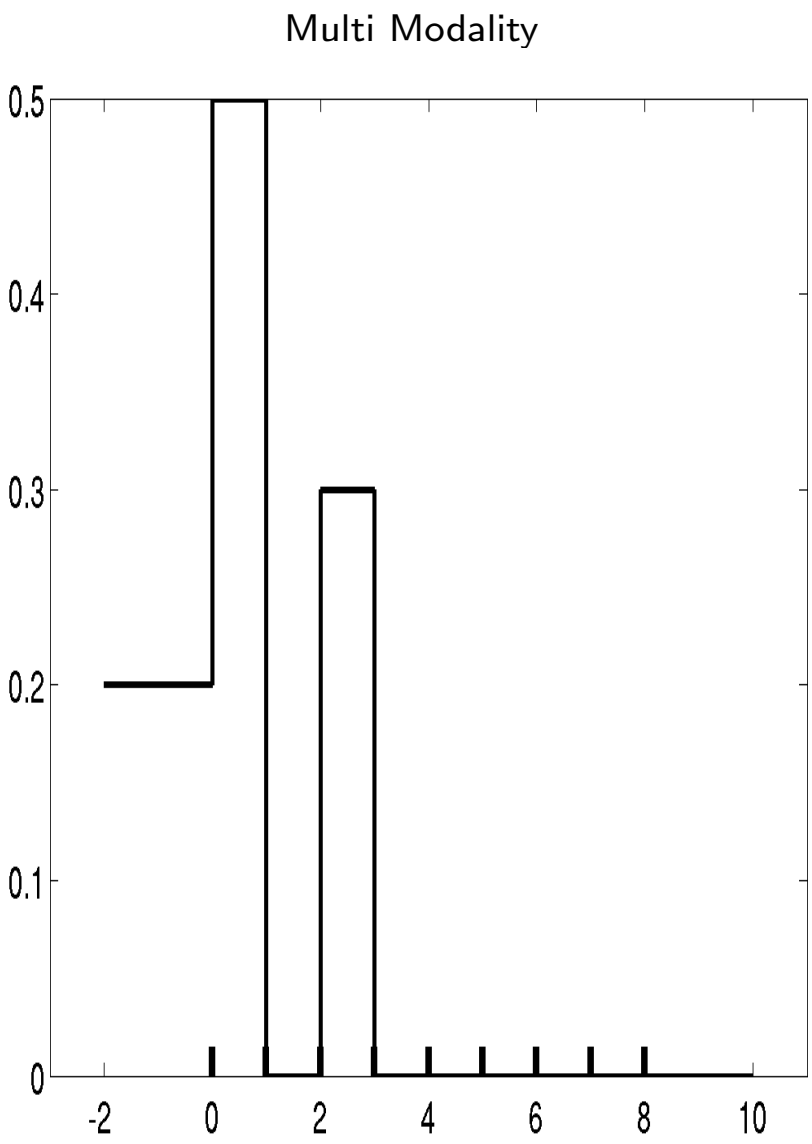
Motivation

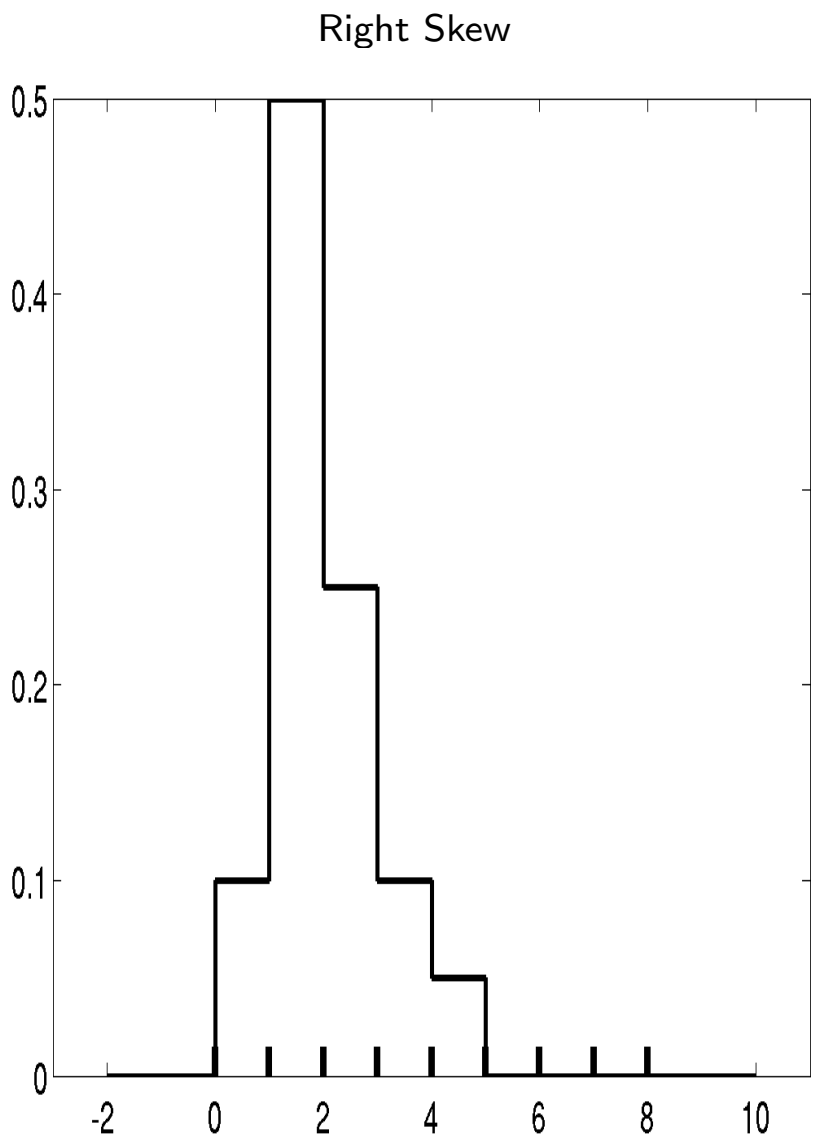
- **Probabilistic surveys** provide a wealth of information beyond point projections (Manski, 2004)
  - "... in a world characterized by pervasive uncertainty, density forecasts provide a comprehensive representation of respondents' views about possible future outcomes for the variables of interest." (Potter, 2016)
  - More and more probabilistic surveys out there: Survey of Consumer Expectations, Survey of Primary Dealers/Market Particpants, ...

The Inference Problem

- Survey respondents only provide **a few points of the CDF** of the predictive distribution: the percent chance that the variable of interest would fall within different pre-specified contiguous ranges/bins

Examples of Individual Distributions from the 2012H2Q1 SPF Survey





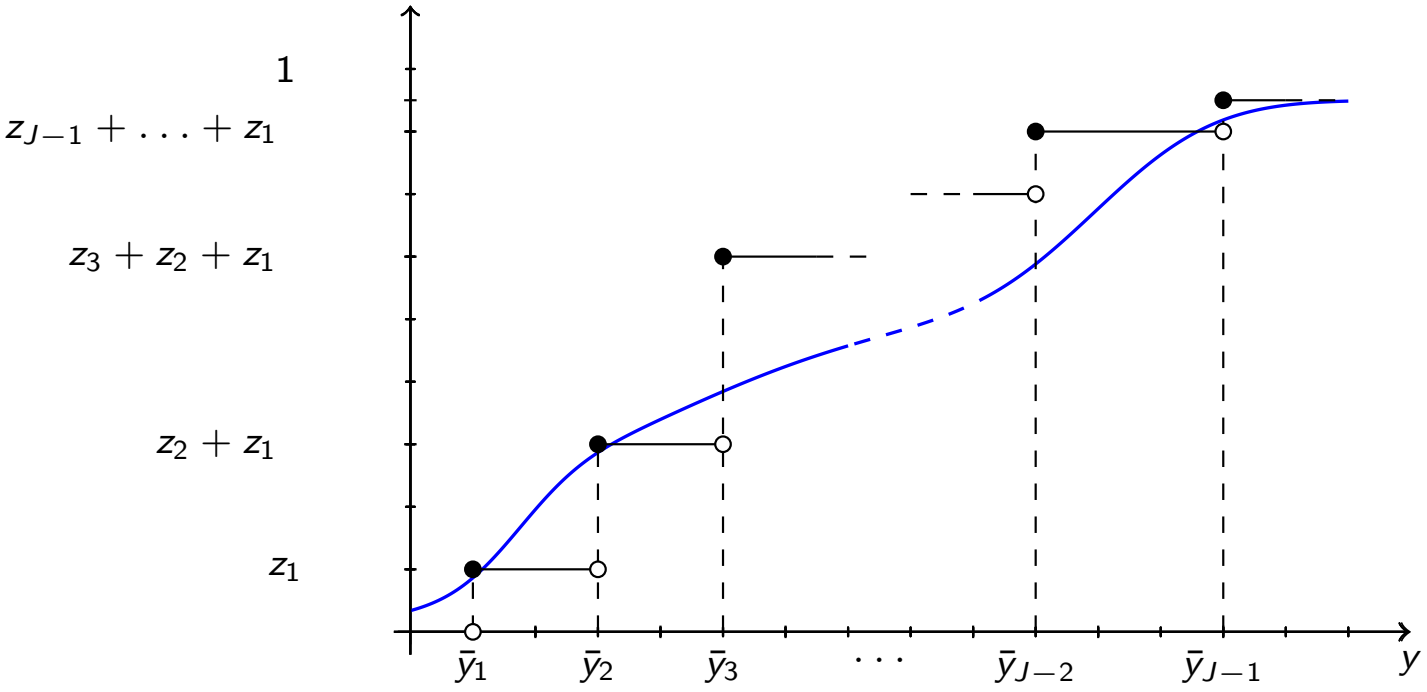
The Inference Problem

- Survey respondents only provide **a few points of the CDF** of the predictive distribution: the percent chance that the variable of interest would fall within different pre-specified contiguous ranges/bins

- *Econometrician's problem:* Use those few points to **conduct inference on the entire subjective predictive density forecast**, and address questions of interest
  - e.g., What is the variance of the aggregate/individual distribution(s)? How did it evolve over time? What is the likelihood of a "return to the seventies", or of deflation? ...

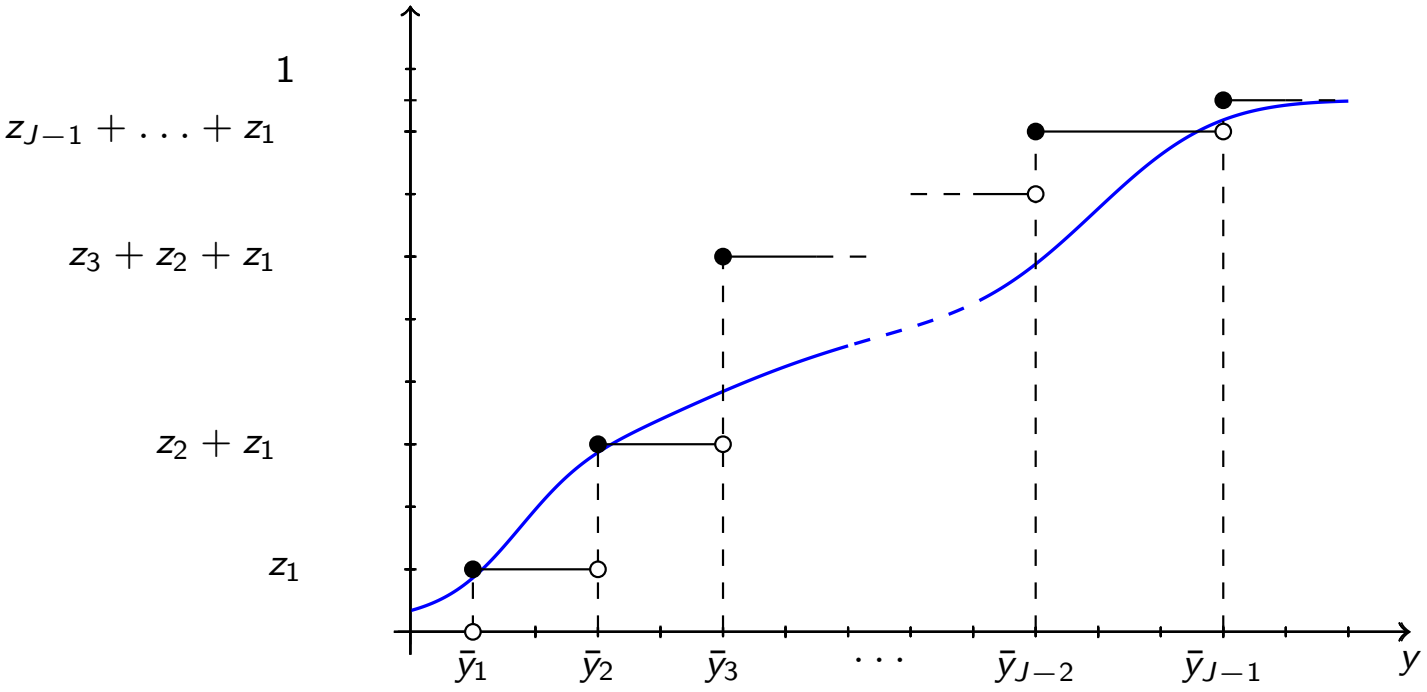
The Inference Problem

- $F(y|\boldsymbol{\theta})$ is the *continuous* subjective distribution with $y \in \mathbb{R}$ on which I am conducting inference
- The $\mathbf{z} = (z_1, \ldots, z_J)$, $\mathbf{z} \in \Delta^J$ are the *reported probabilities* on bins
- ... which are the probability implied by the $F(y|\boldsymbol{\theta})$ plus some "noise" (approximation, misreporting,...)



The Inference Problem

- $F(y|\boldsymbol{\theta})$ is the *continuous* subjective distribution with $y \in \mathbb{R}$ on which I am conducting inference
- The $\mathbf{z} = (z_1, \ldots, z_J)$, $\mathbf{z} \in \Delta^J$ are the *reported probabilities* on bins
- ... which are the probability implied by the $F(y|\boldsymbol{\theta})$ plus some "noise" (approximation, misreporting,...)



- **Inference uncertainty** arises from 1) *noise*, 2) *lack of identification*

Approaches So Far

- Literature so far has <u>not</u> (to our knowledge) treated it as an **inference problem**, but just fit <u>one</u> parametric distribution on the CDF points. That is, pick an $\tilde{F}(.|\tilde{\theta})$ and find

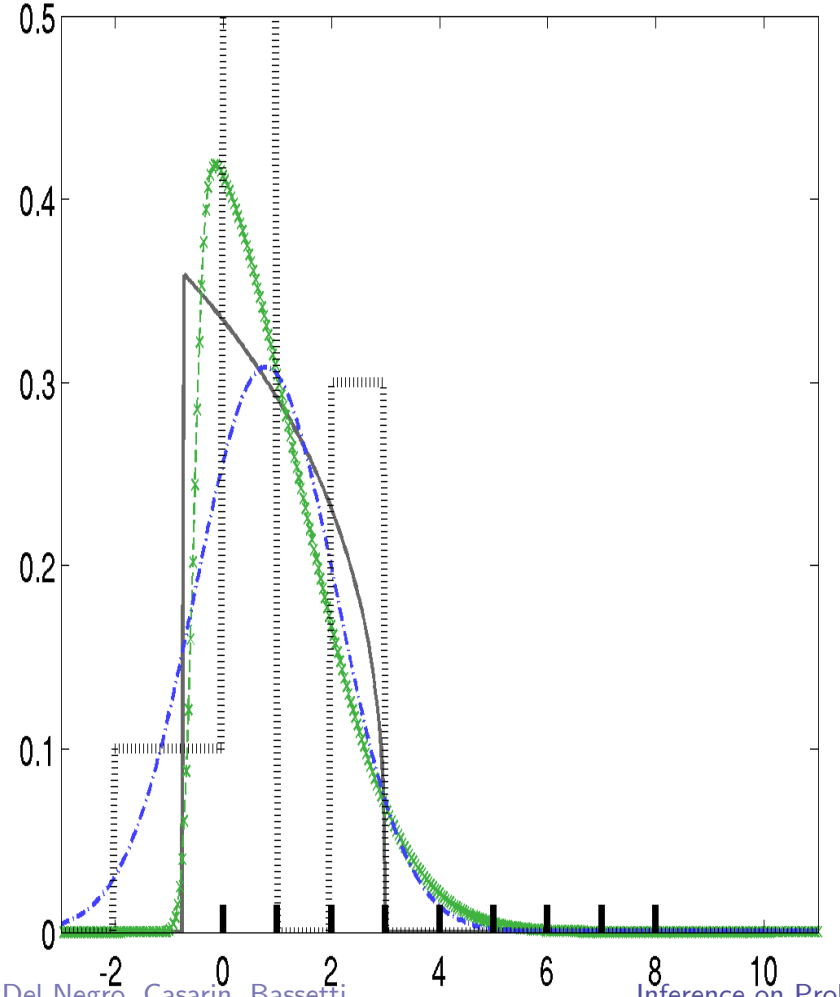
$$\tilde{\theta}_* = argmin_{\tilde{\theta}} \sum_{j=1}^{J} \left(\sum_{s=1}^{j} z_s - \tilde{F}(y_j|\tilde{\theta}) \right)$$

- ... *without characterizing inference uncertainty*

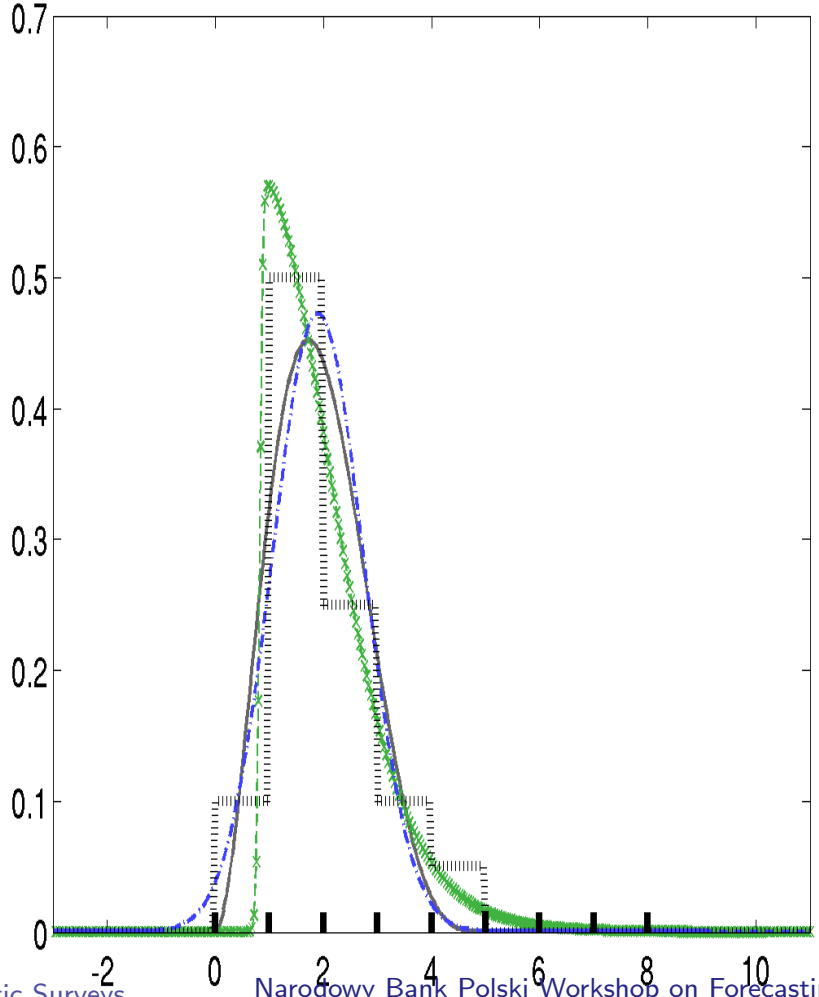
Approaches So Far

- *Normal*: Giordani and Söderlind, 2003, Clements, 2014, ...; *Beta*: Engelberg, Manski and Williams, 2009; *Skew-t*: Ganics et al, 2018, ...

Multi Modality



Right Skew



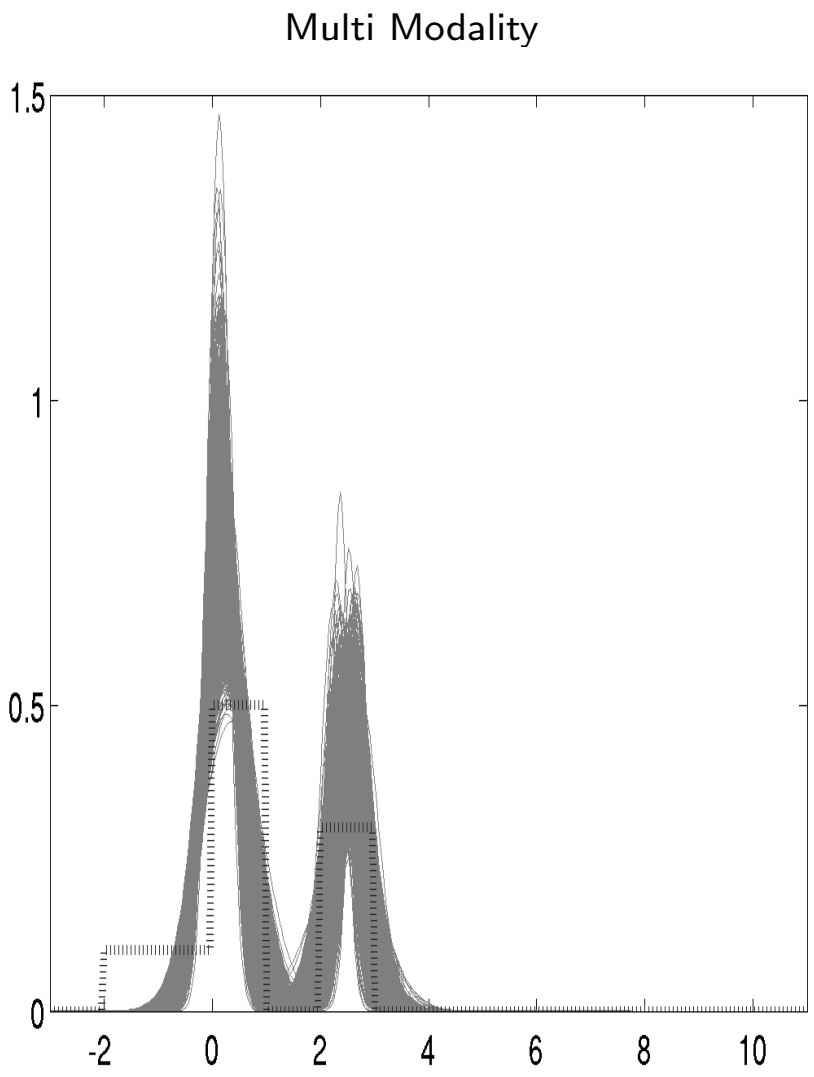
Methodological Contribution

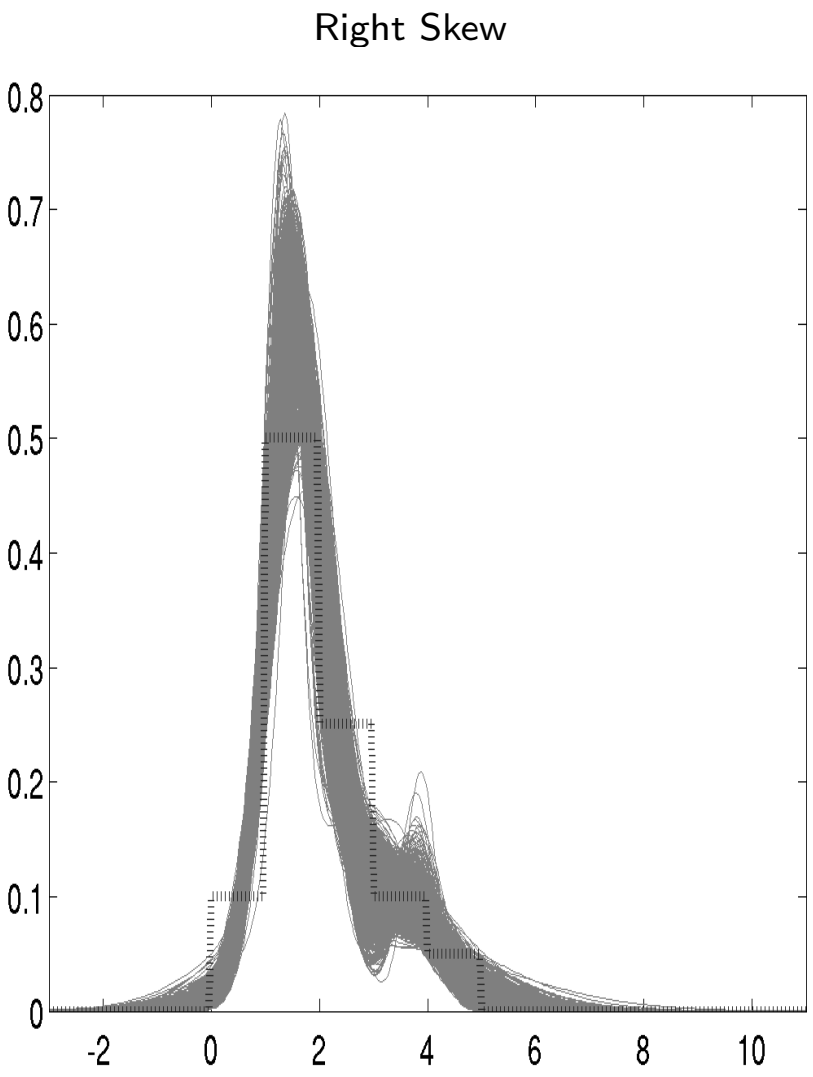
1. Kernel distribution
   - Underlying continuous distribution is *mixture of two Gaussians* as opposed to (one) Gaussian, or Beta, or Skew-t ...
   - **Model "noise" /rounding to zero**

2. We provide an explicit probability model for the data $\rightarrow$ **inference**

3. We use a **Bayesian non-parametric approach**: characterize both density forecasts and the "noise" in a *flexible* way
   - Use (potentially infinite) mixture of the kernel distributions
   - (Data-dependent) *pooling* $\rightarrow$ fewer parameters to estimate

Our Approach

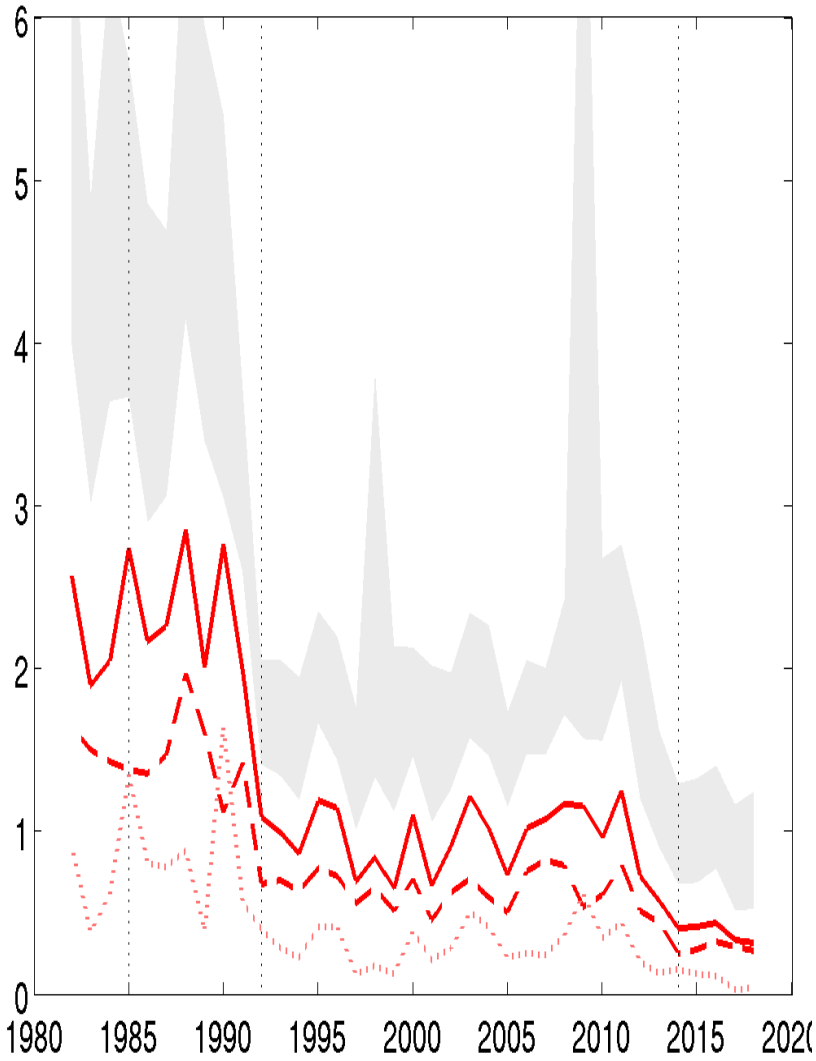




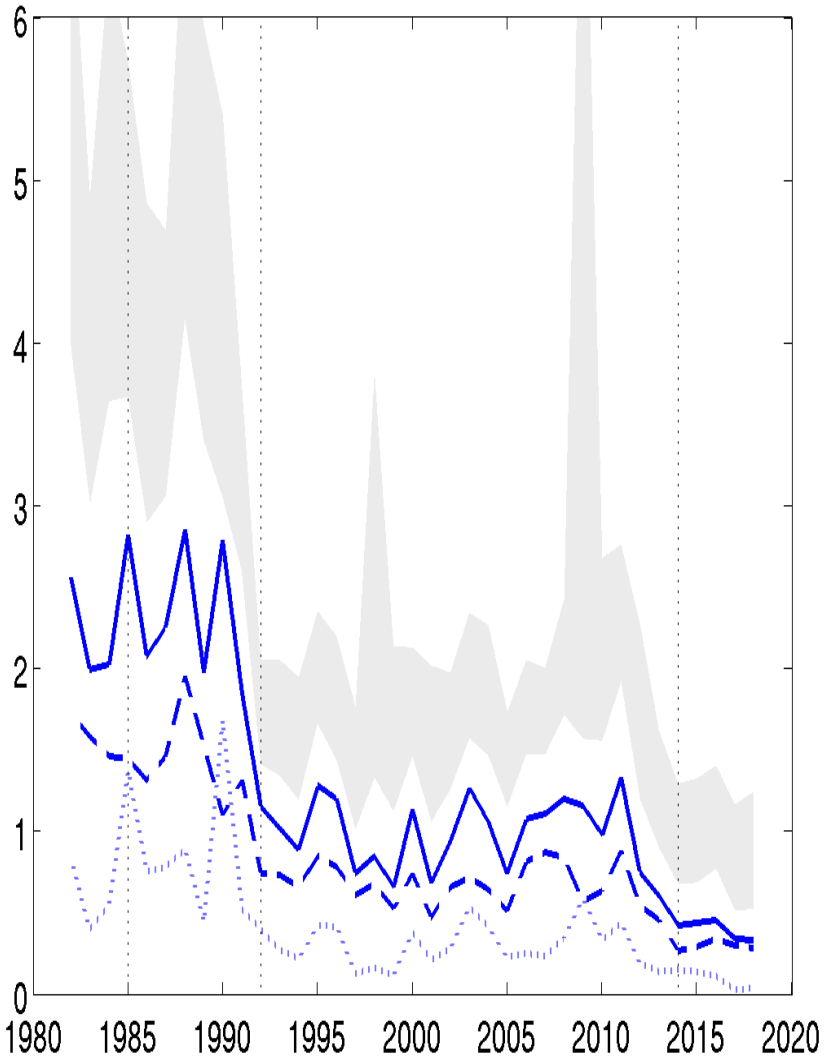
... And It Makes a Difference

Variance, Uncertainty, and Disagreement for Inflation (H2Q2)

Beta



Normal



The Data

- y is the variable being forecasted (in our application, year-over-year GDP deflator inflation or real output growth over the following year).

- $\bar{y}_1 < \ldots < \bar{y}_J$ is the set of *bin upper bounds*, with $\bar{y}_J$ possibly being $+\infty$ (lowest bin also open). Bins are mutually exclusive, contiguous, and are chosen by the survey designer. (In our application they cover the entire real line).

- For each forecaster $i = 1, ..., n$, the 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 falls within the respective bin. In our application, the z_i's are from the Survey of Professional Forecasters.

- Many z_{ij}'s are zero. ξ_i is a vector of indicator functions: $\xi_j = 0$ if and only if $z_j = 0$ ($\xi_j = 1$ otherwise).

Outline of Model Description

- We model $\mathbf{z}_i$'s using a **Bayesian non-parametric model**. Specifically, each $\mathbf{z}_i$ is independently distributed according to a (potentially) infinite mixture of kernel distributions $h(\mathbf{z}_i|\boldsymbol{\theta})$

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

- We first describe the kernel $h(\mathbf{z}_i|\boldsymbol{\theta})$ and then discuss the Bayesian Non-Parametrics

Outline of Model Description

- We model $\mathbf{z}_i$'s using a **Bayesian non-parametric model**. Specifically, each $\mathbf{z}_i$ is independently distributed according to a (potentially) infinite mixture of kernel distributions $h(\mathbf{z}_i|\boldsymbol{\theta})$

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

- We first describe the kernel $h(\mathbf{z}_i|\boldsymbol{\theta})$ and then discuss the Bayesian Non-Parametrics

- The kernel itself is given by a **hierarchical model**
  - For each forecaster there is an underlying continuous probability distribution $F_i(y)$ that determines the probabilities ν_i associated with the bins.
  - The z_i's are a "noisy" version of the ν_i's

Kernel Distribution: Noise

- **Challenge 1**: Modeling the "**noise**" – the difference between the reported probabilities associated with each bin and the true underlying probabilities that forecasters have in mind (e.g., reporting mistake, rounding, ...) $\rightarrow$ Dirichlet distribution

- <u>If</u> (relaxed on the next slide) $z_j > 0$ for all $j = 1, \ldots, J$ (i.e. **no zeros**) and $z_1 + \ldots + z_J = 1$, then $\mathbf{z} = (z_1, \ldots, z_J)$ follows a standard **Dirichlet distribution**:

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

  - $\nu_j(\boldsymbol{\theta})$: the underlying unknown probability on the jth bin (implied by the underlying $F(y|\boldsymbol{\theta})$).
  - $\phi(\boldsymbol{\theta})$: the unknown precision level.

Kernel Distribution: Noise—continued

- **Challenge 2**: Modeling the "**probability of reporting zeros**" – (e.g., forecasters assign zero probability to a bin if the true probability is small enough) $\rightarrow$ zero-adjusted Dirichlet model

- $\mathbf{z} = (z_1, \ldots, z_J)$ and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_J)$ follows a **zero-adjusted Dirichlet distribution**

$$h(\mathbf{z}, \boldsymbol{\xi} | \boldsymbol{\theta}) = \prod_{j=1}^{J} \underbrace{\alpha_j(\boldsymbol{\theta})^{\xi_j}}_{\text{Probability of } z_j = 0} \cdot \underbrace{(1 - \alpha_j(\boldsymbol{\theta}))^{1-\xi_j} \tilde{h}(\tilde{\mathbf{z}} | \boldsymbol{\theta}, \phi, \boldsymbol{\xi})}_{\text{Probability of } z_j \neq 0}$$

- $\tilde{\mathbf{z}}$—the set of non-zero observed probabilities $\{z_j > 0\}$—follows a "renormalized" Dirichlet

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

where $\tilde{\mathcal{J}} = \{j = 1, \ldots, J : z_j > 0\}$ and $\tilde{\nu}_j(\boldsymbol{\theta}) = \dfrac{\nu_j(\boldsymbol{\theta})}{\sum_{j \in \tilde{\mathcal{J}}} \nu_j(\boldsymbol{\theta})}$ are the renormalized ν's for $j \in \tilde{\mathcal{J}}$

Kernel Distribution: Probability of Reporting 0's

- α_j: the probability that a forecaster will report zero on the jth bin
- **In our application the probability of zero-valued observations** is modeled as the cdf of a beta distribution $b(.|\nu, r)$ with mean ν and precision r

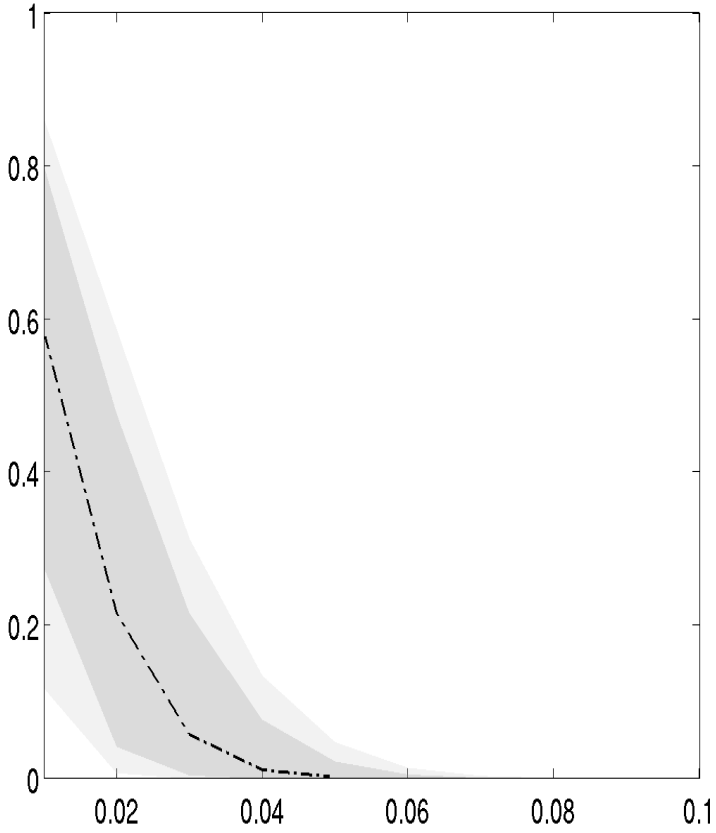
$$\alpha_j(\boldsymbol{\theta}) = \int_0^{\varepsilon} b(x|\nu_j(\boldsymbol{\theta}), r)dx$$

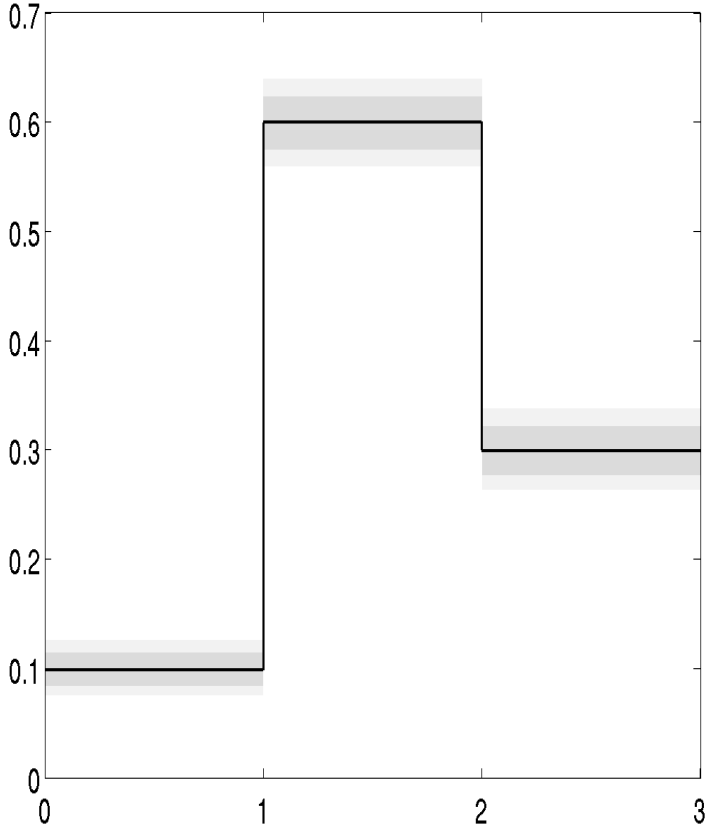
Probability of Forecaster Reporting Zero Mass and Noise

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

Noise

as a function of ν





Kernel Distribution: the $\nu(\boldsymbol{\theta})$'s

- Forecasters have an **underlying subjective probability** distribution for y, whose cdf is $F(y|\boldsymbol{\theta})$, associated with a vector of probabilities over the J bins:

$$\nu_j(\boldsymbol{\theta}) = F(\bar{y}_j|\boldsymbol{\theta}) - F(\bar{y}_{j-1}|\boldsymbol{\theta}),\ j = 1, \ldots, J,$$

where $\nu_j \geq 0$, $\sum_{j=1}^{J} \nu_j = 1$ $(\bar{y}_0 = -\infty)$

  - **In our application $F(\cdot|\theta)$ is a mixture of two Gaussian distributions**

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

Bayesian Non-Parametrics: Dirichlet Process Prior

- We model $\mathbf{z}_i$'s using a **Bayesian non-parametric model**. Specifically, each $\mathbf{z}_i$ is independently distributed according to

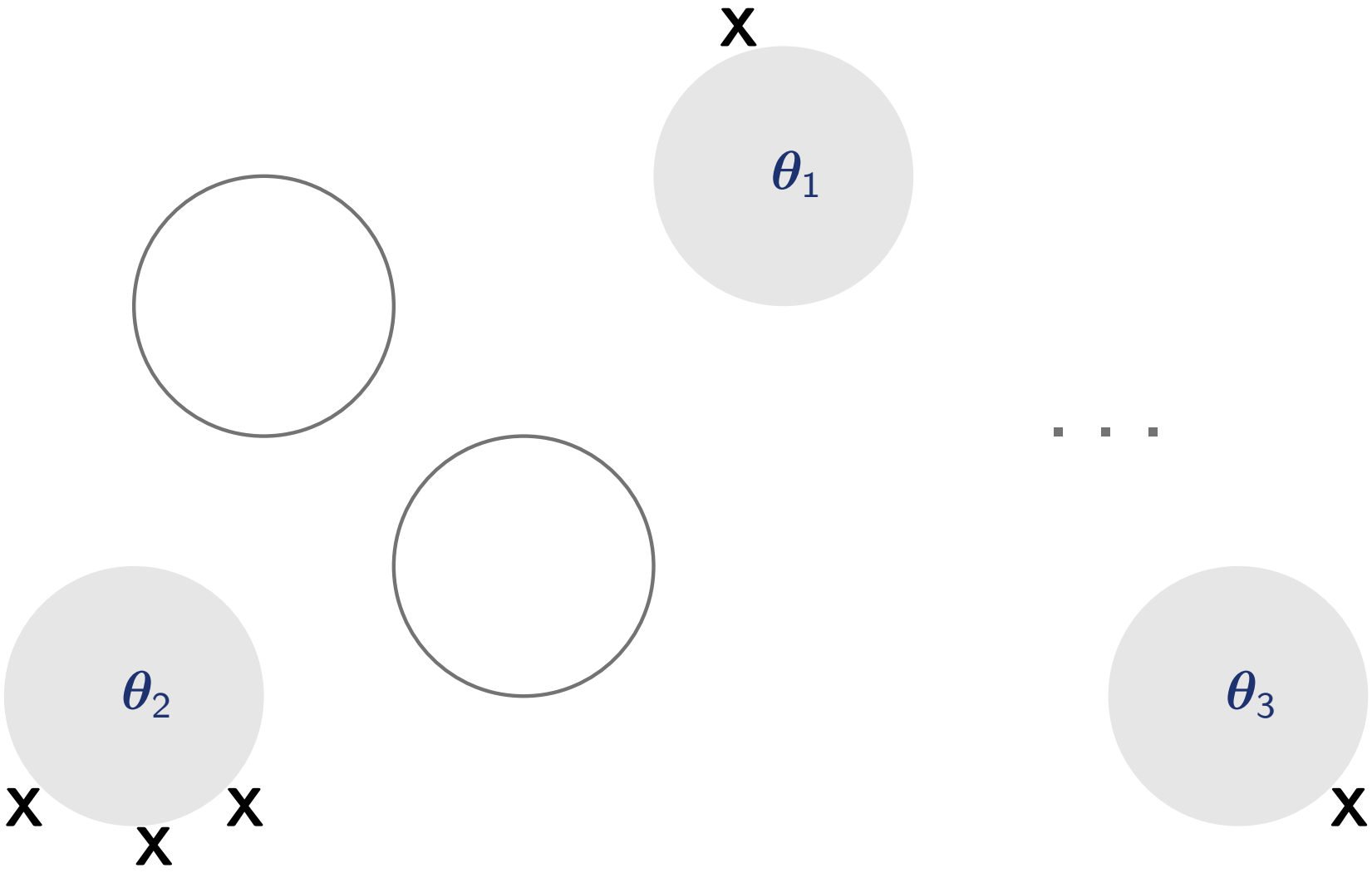
$$h(\mathbf{z}_i|G) = \int h(\mathbf{z}_i|\boldsymbol{\theta})G(d\boldsymbol{\theta})$$

where G is a **Dirichlet process** $\mathcal{DP}(\psi, G_0)$, which is equivalent to the Bayesian hierarchical model

$$\begin{aligned}
\mathbf{z}_i|G &\overset{ind}{\sim} h(\mathbf{z}|\boldsymbol{\theta}_i), \quad i = 1, \ldots, n \\
\boldsymbol{\theta}_i|G &\overset{iid}{\sim} G(\boldsymbol{\theta}) \\
G &\sim \mathcal{DP}(\psi, G_0)
\end{aligned}$$

and $G_0(\boldsymbol{\theta})$ is the base measure and ψ the concentration parameter.

Bayesian Non-Parametrics: Chinese Restaurant Process



Pólya's urn: $\boldsymbol{\theta}_n | \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{n-1} \sim \frac{\psi}{\psi + n - 1} G_0(\boldsymbol{\theta}_n) + \frac{1}{\psi + n - 1} \sum_{i=1}^{n-1} \delta_{\boldsymbol{\theta}_i}(d\boldsymbol{\theta})$

Bayesian Non-Parametrics: Infinite Mixture Representation

- G can be represented as a **discrete random measure** (Sethuraman, 1994)

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

  where the "atoms" $\boldsymbol{\theta}_k$ are i.i.d. from the base measure G_0 (prior), and the random weights w_k are generated by the stick-breaking construction $w_k = v_k \prod_{l=1}^{k-1}(1 - v_l)$, where the v_l's are drawn i.i.d. from a Beta distribution $\mathcal{B}e(1, \psi)$

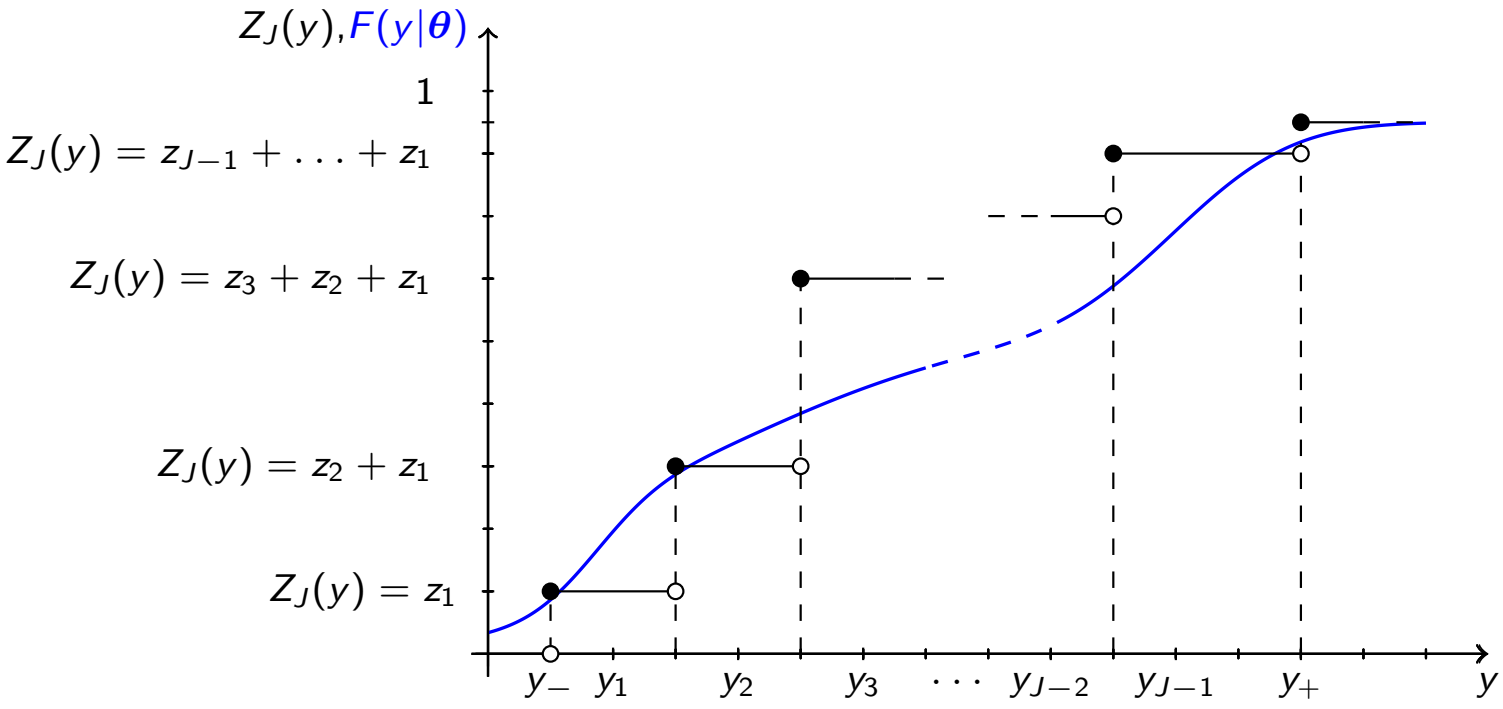
- The model has the **infinite** *mixture representation*:

$$h(\mathbf{z}|w_{1:\infty}, \boldsymbol{\theta}_{1:\infty}) = \sum_{k=1}^{\infty} w_k h(\mathbf{z}|\boldsymbol{\theta}_k)$$

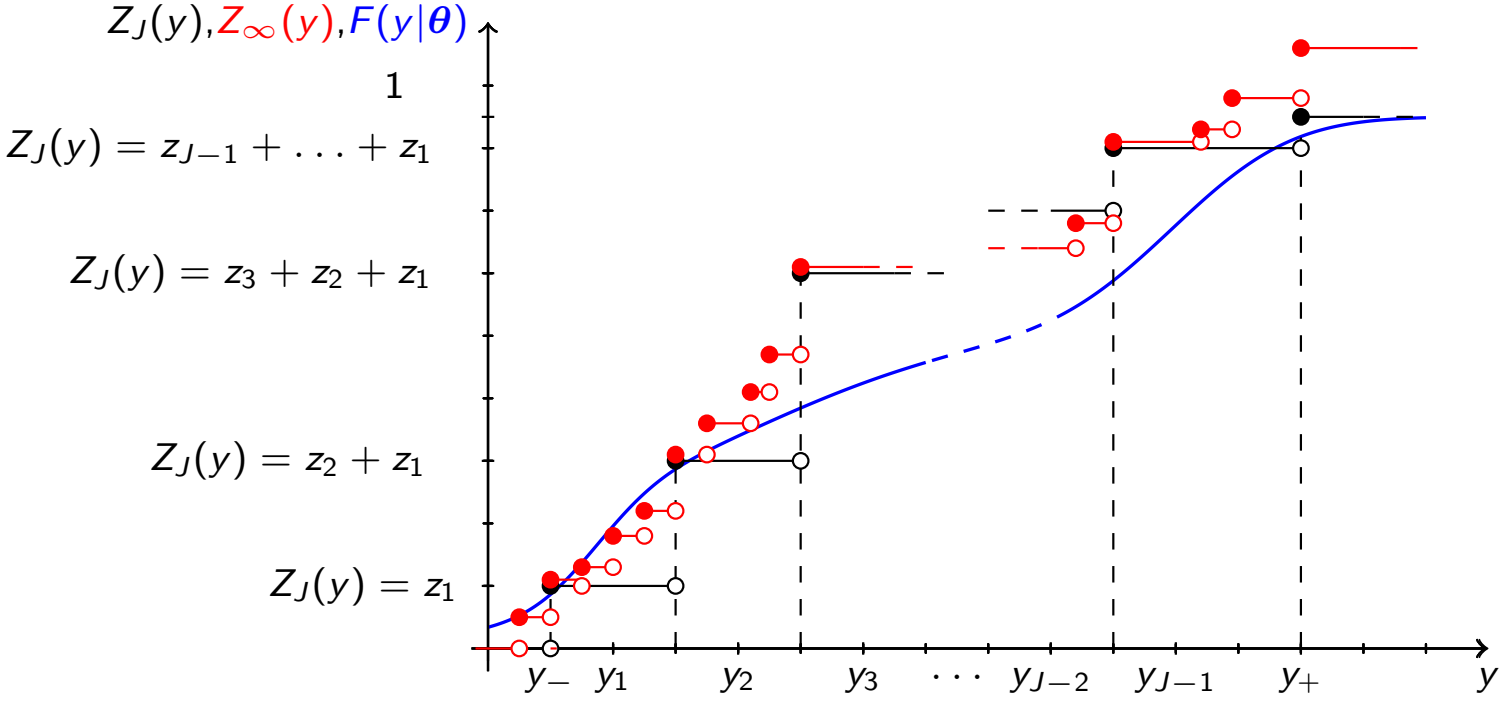
Properties of the Kernel: A Random Histogram Model

Relationship between:

- histogram probabilities $\mathbf{z} = (z_1, \ldots, z_J) \sim h(\mathbf{z}|\boldsymbol{\theta})$, $\mathbf{z} \in \Delta^J$
- the subjective distribution $F(y|\boldsymbol{\theta})$ with $y \in \mathbb{R}$ inflation
- and the cumulative histogram probabilities $Z_J(y) \in [0, 1]$



The Limiting Random Histogram



As the number of <u>bins</u> goes to infinity ($J \to \infty$):

- the cdf $Z_J(y)$ converges to the random cdf $Z_\infty(y)$
- $\lim_{J\to\infty} \mathbb{E}(Z_J(y)|\boldsymbol{\theta}) = \mathbb{E}(Z_\infty(y)|\boldsymbol{\theta}) = F(y|\boldsymbol{\theta})$

Model Properties

- So far this was for given $\boldsymbol{\theta}_i$. BNP introduces flexibility over $F(y|\boldsymbol{\theta}_i)$:

$$\boldsymbol{\theta}_i \overset{i.i.d.}{\sim} G, \quad G \sim \mathcal{DP}(\psi, G_0).$$

- $\int F(y|\boldsymbol{\theta}_i)G(d\boldsymbol{\theta}_i)$ can approximate *any* continuous distribution $H(y)$
- $P(\{Z_\infty \in U_H\}) > 0$ for any weak neighbourhood U_H of $H(y)$: loosely speaking, **even if** $H(y)$ **is not** $F(y|\boldsymbol{\theta}_i)$ (e.g., a mixture of two Gaussians), **a priori our model puts positive probability on it**.

Model Properties

- So far this was for given $\boldsymbol{\theta}_i$. BNP introduces flexibility over $F(y|\boldsymbol{\theta}_i)$:

$$\boldsymbol{\theta}_i \overset{i.i.d.}{\sim} G, \quad G \sim \mathcal{DP}(\psi, G_0).$$

- $\int F(y|\boldsymbol{\theta}_i)G(d\boldsymbol{\theta}_i)$ can approximate *any* continuous distribution $H(y)$
- $P(\{Z_\infty \in U_H\}) > 0$ for any weak neighbourhood U_H of $H(y)$: loosely speaking, **even if** $H(y)$ **is not** $F(y|\boldsymbol{\theta}_i)$ (e.g., a mixture of two Gaussians), **a priori our model puts positive probability on it**.
- **Identification**. The $\nu(\boldsymbol{\theta})$'s (points in the CDF corresponding to the bins) are identified, whereas $F(\cdot|\boldsymbol{\theta})$ is not. The prior information marginalized to $\nu(\boldsymbol{\theta})$ is the *revisable prior knowledge* and the conditional prior of $F(\cdot|\boldsymbol{\theta})$ given $\nu(\boldsymbol{\theta})$ is the *unrevisable prior knowledge* [Kitagawa (2012), Moon and Schorfheide (2011), Florens and Simoni (2012)].
- A priori our model places positive probability on *any* $H(y)$ going through these points.

Asymptotic Properties: Posterior Consistency

- Let $\Pi(h(\cdot))$ and $\Pi(h(\cdot)|\mathbf{z}_{1:n})$ be the **prior** and **posterior distribution**, respectively —viewed as *distributions over the densities* $h(\mathbf{z})$ *on the simplex* Δ^J (the bin probabilities $\mathbf{z} = (z_1, \ldots, z_J)$).
- Let $h_0(\mathbf{z})$ be the DGP (*note that our model can be misspecified*)

Asymptotic Properties: Posterior Consistency

- Let $\Pi(h(\cdot))$ and $\Pi(h(\cdot)|\mathbf{z}_{1:n})$ be the **prior** and **posterior distribution**, respectively —viewed as *distributions over the densities* $h(\mathbf{z})$ *on the simplex* Δ^J (the bin probabilities $\mathbf{z} = (z_1, \ldots, z_J)$).
- Let $h_0(\mathbf{z})$ be the DGP (*note that our model can be misspecified*)
- The **posterior distribution** $\Pi(\cdot|\mathbf{z}_{1:n})$ on the set of distributions/densities is weakly consistent at h_0 in the sense that as $n \to \infty$ it **converges** to a Dirac centered at the true distribution h_0 a.s. ($\Pi(\cdot|\mathbf{z}_{1:n}) \to \delta_{h_0}(\cdot)$):

$$\Pi(h : d_w(h_n, h_0) \leq \varepsilon|\mathbf{z}_{1:n}) \xrightarrow[n \longrightarrow \infty]{a.s.} 1, \quad \forall \varepsilon > 0$$

where d_w is a metric of the weak convergence topology (e.g. the Lévy-Prokhorov metric) on the set of measures/densities.

- Note: we say that $d_w(h_n, h_0) \to 0$ and write $h_n \xrightarrow{w} h_0$, if

$$\left| \int g(\mathbf{z}) h_n(\mathbf{z}) d\mathbf{z} - \int g(\mathbf{z}) h_0(\mathbf{z}) d\mathbf{z} \right| \longrightarrow 0$$

as $n \to \infty$, for every bounded continuous function g.

Asymptotic Properties: Posterior consistency

- Even if the model is misspecified, we still have consistency.
- If $F(y|\boldsymbol{\theta})$ is flexible enough that the ν's can assume any value (hence they are correctly specified), then an infinite mixture of Dirchlet can approximate any distribution on the simplex $\rightarrow$ weak posterior consistency holds for *any* $h_0(\mathbf{z})$ (e.g., different model for rounding).

Asymptotic Properties: Posterior consistency

- Even if the model is misspecified, we still have consistency.
- If $F(y|\boldsymbol{\theta})$ is flexible enough that the ν's can assume any value (hence they are correctly specified), then an infinite mixture of Dirchlet can approximate any distribution on the simplex $\rightarrow$ weak posterior consistency holds for *any* $h_0(\mathbf{z})$ (e.g., different model for rounding).
- If $h_0(\mathbf{z})$ can be approximated by an *infinite* mixture of our kernels $h(\mathbf{z}|\boldsymbol{\theta})$, posterior consistency applies.

Summary of Asymptotic Properties

1. **Prior**
   - BNP flexibility implies that a priori our model places positive probability on *any* $H(y)$
   - As the number of <u>bins</u> goes to infinity, the cumulated histogram probabilities $Z_J(y)$ converge to a **latent random distribution** $Z_\infty(y)$ whose mean is the subjective underlying continuous distribution $H(y)$

Summary of Asymptotic Properties

1. **Prior**
   - BNP flexibility implies that a priori our model places positive probability on *any* $H(y)$
   - As the number of bins goes to infinity, the cumulated histogram probabilities $Z_J(y)$ converge to a **latent random distribution** $Z_\infty(y)$ whose mean is the subjective underlying continuous distribution $H(y)$

2. **Posterior**: for given bins, as the number of forecasters (observations) goes to infinity, the **posterior distribution** —viewed as the distribution over the probabilities associated with bins, that is, the *histogram* (not the underlying continuous distribution) —is weakly **consistent** (arbitrarily close in the weak topology of measures to the true distribution)

Sampling: Finite Mixture Representation

- Data augmentation:

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

where $u_i \sim U[0,1]$ ("slicing": Walker, 2007; Kalli et al., 2011)

- Likelihood:

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

where the d_i's are auxiliary indicators equal to k if we draw from the kth mixture component ($d_i \in \{k : u_i < w_k\}$)

Sampling: Posterior Approximation

- Each forecaster i is described by a mixture

$$\begin{aligned} F_i(y) &= \mathbb{E}\left[F(y|\boldsymbol{\theta}_{d_i})|\mathbf{z}_{1:n}, \boldsymbol{\xi}_{1:n}\right] \\ &= \mathbb{E}\left[\sum_{k=1}^{\infty} F(y|\boldsymbol{\theta}_k)\mathbb{I}\{d_i = k\}\middle|\mathbf{z}_{1:n}, \boldsymbol{\xi}_{1:n}\right] \end{aligned}$$

- The output of the Gibbs sampler can be used to approximate it as follows

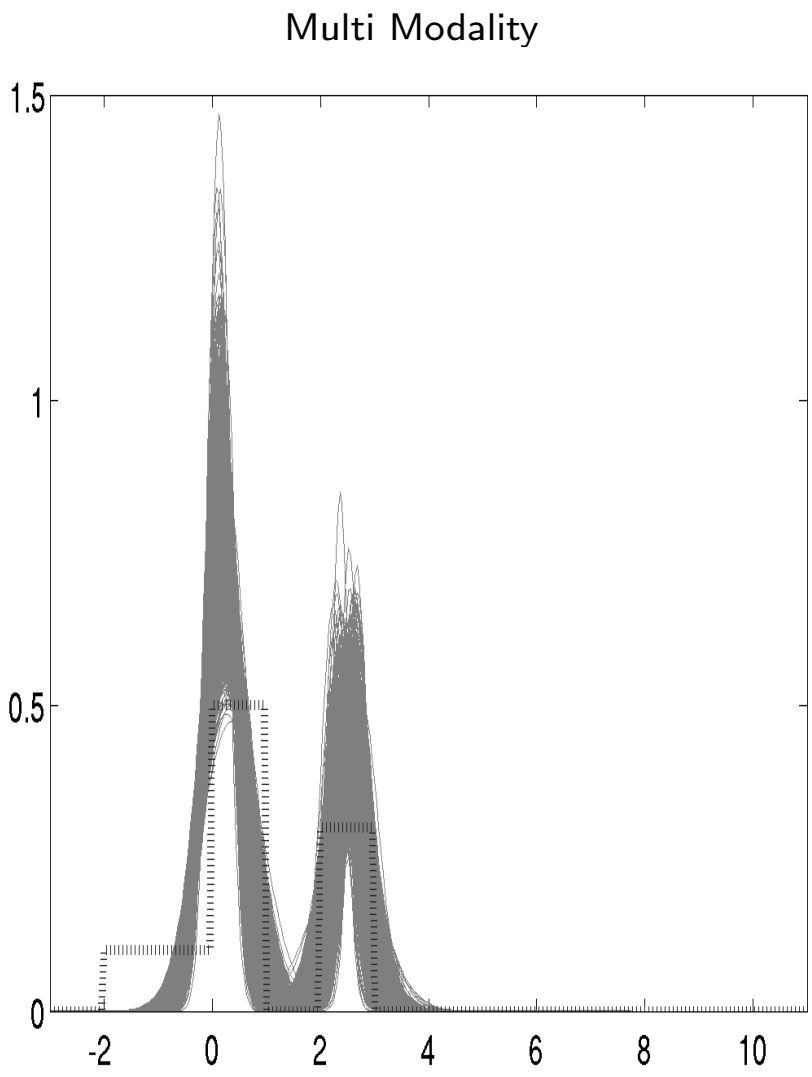
$$\hat{F}_i(y) = \frac{1}{M}\sum_{m=1}^{M}\sum_{k=1}^{\infty} F(y|\boldsymbol{\theta}_k^{(m)})\mathbb{I}\{d_i^{(m)} = k\}$$

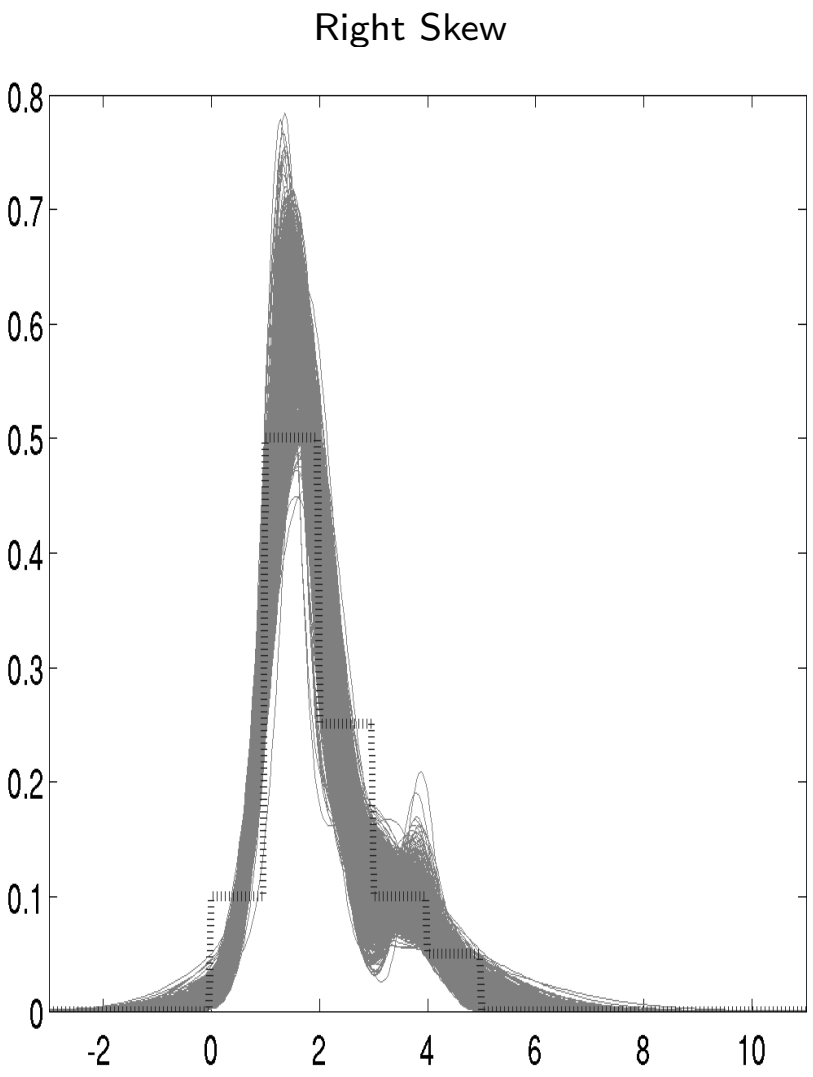
where $m = 1, \ldots, M$ are the MCMC samples for the infinite mixture atoms and allocation variables.

Base Measure and Hyper-Priors

- **The base measure G_0 is the same for all years**, and is the same for output growth and inflation.
- The hyper-priors in the base measure are independent across parameters:
- $\mu \sim \mathcal{N}(2, 5^2)$
- $\sigma_1 \sim \mathcal{IG}a(a_\sigma, b_\sigma)\mathcal{I}(\sigma_1)_{(0,10)}$ where a_σ, b_σ s.t. $E[\sigma_1] = 2,\ V[\sigma_1] = 4$
- $\mu_\delta \sim \mathcal{N}(0, 5^2)$
- $\sigma_2 \sim \mathcal{IG}a(a_\sigma, b_\sigma)\mathcal{I}(\sigma_2)_{(0,10)}$ — same distribution as σ_1.
- $\omega \sim \mathcal{B}(0.5, 3)$
- $\phi \sim \mathcal{G}a(a_\phi, b_\phi)\mathcal{I}(\phi)_{(\underline{\phi},+\infty)}$ where a_ϕ b_ϕ are chosen s.t. the noise has a st. dev. of 2.5% under $E[\phi]$
- $\varepsilon \sim \mathcal{G}a^*(a_\varepsilon, b_\varepsilon)\mathcal{I}(\varepsilon)_{(0,\bar{\varepsilon})}$ where a_ε and b_ε are chosen so that $\int_0^{E[\varepsilon]} b(x|0.025, 100)dx = 0.25$
- $\psi = 1 \rightarrow$ the expected number of clusters for a cross-section of 30 survey respondents is roughly 4

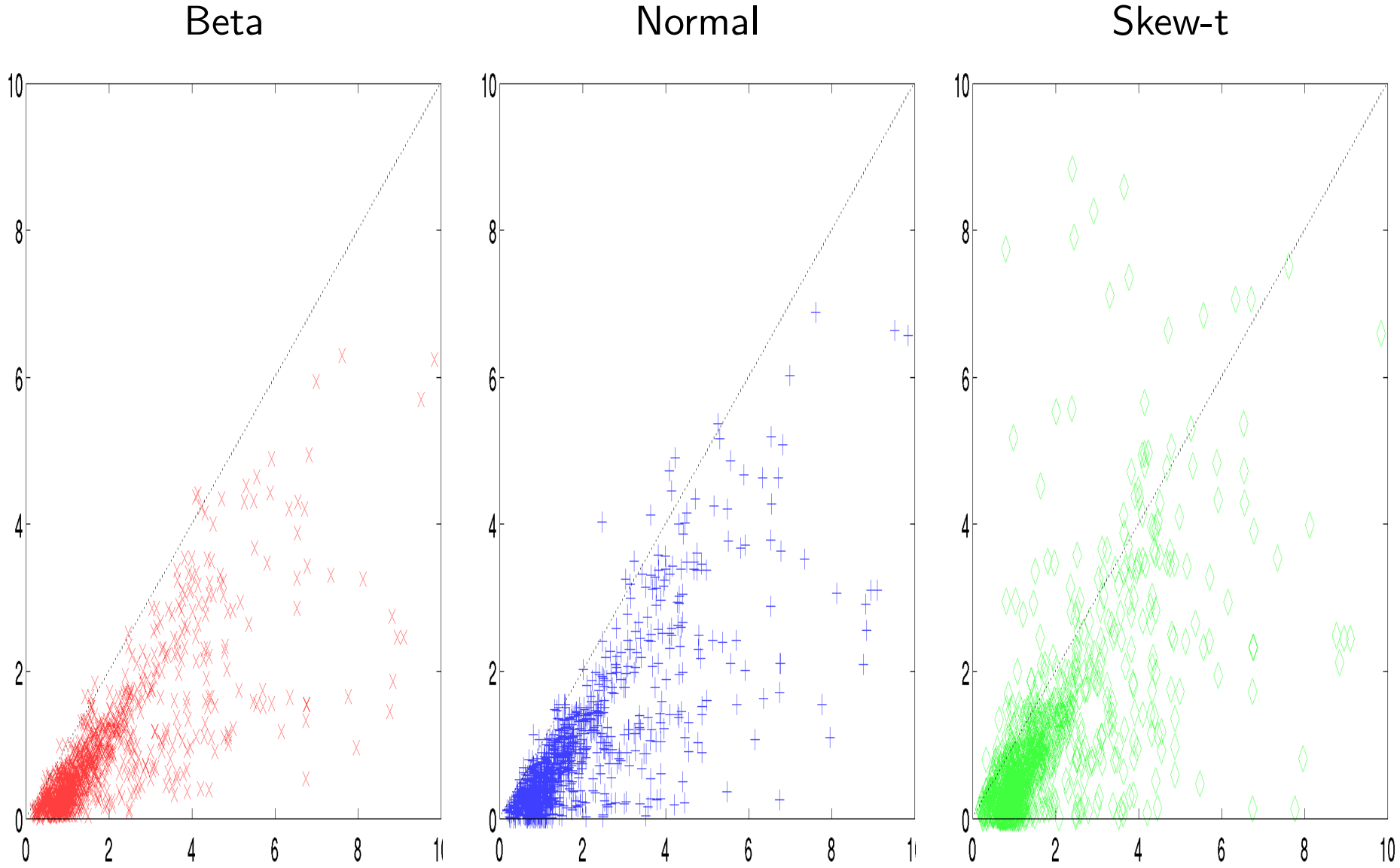
Our Approach





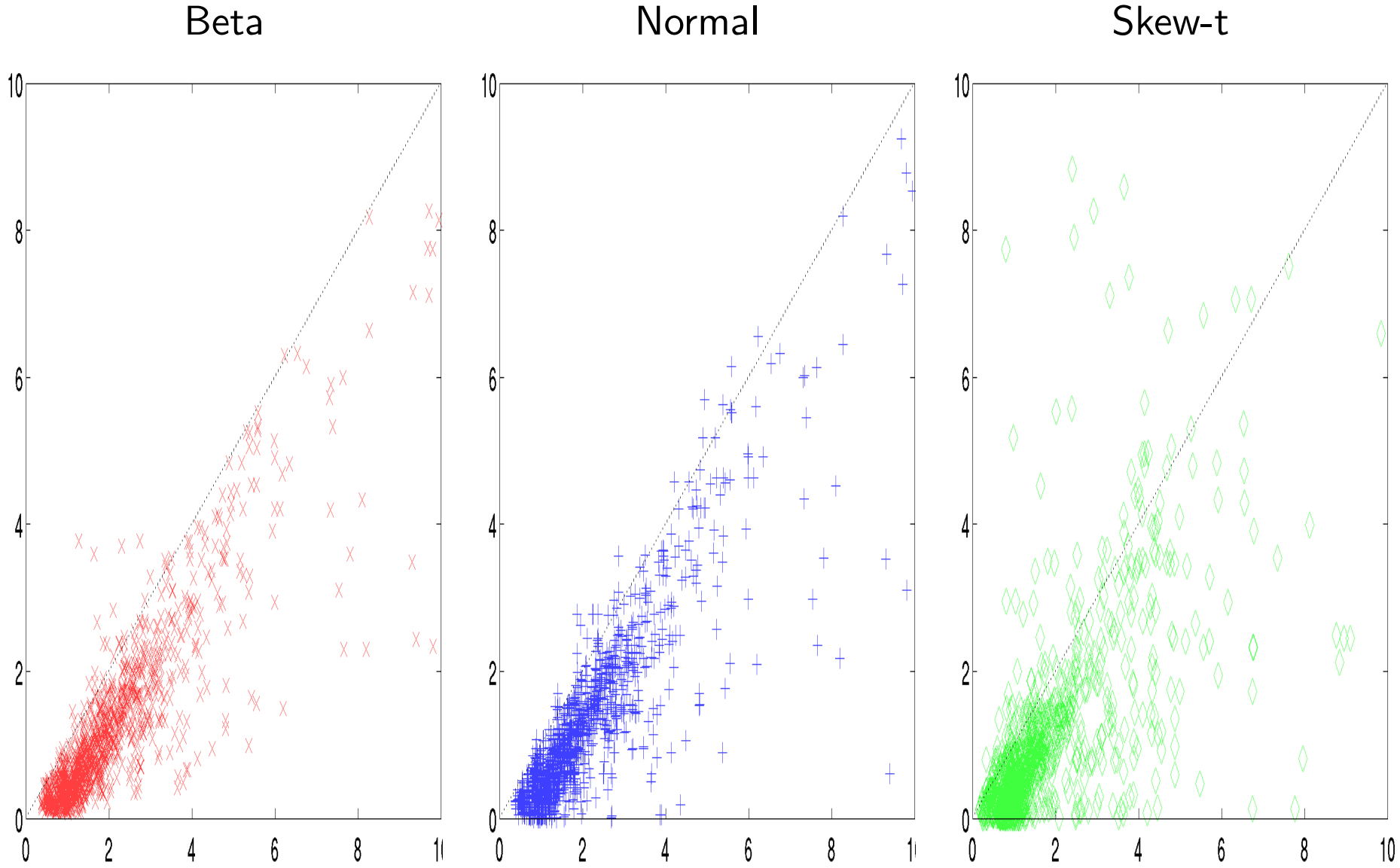
Individual Variances: Our Approach (x-axis) vs ... (y-axis)

Inflation (H2Q2)



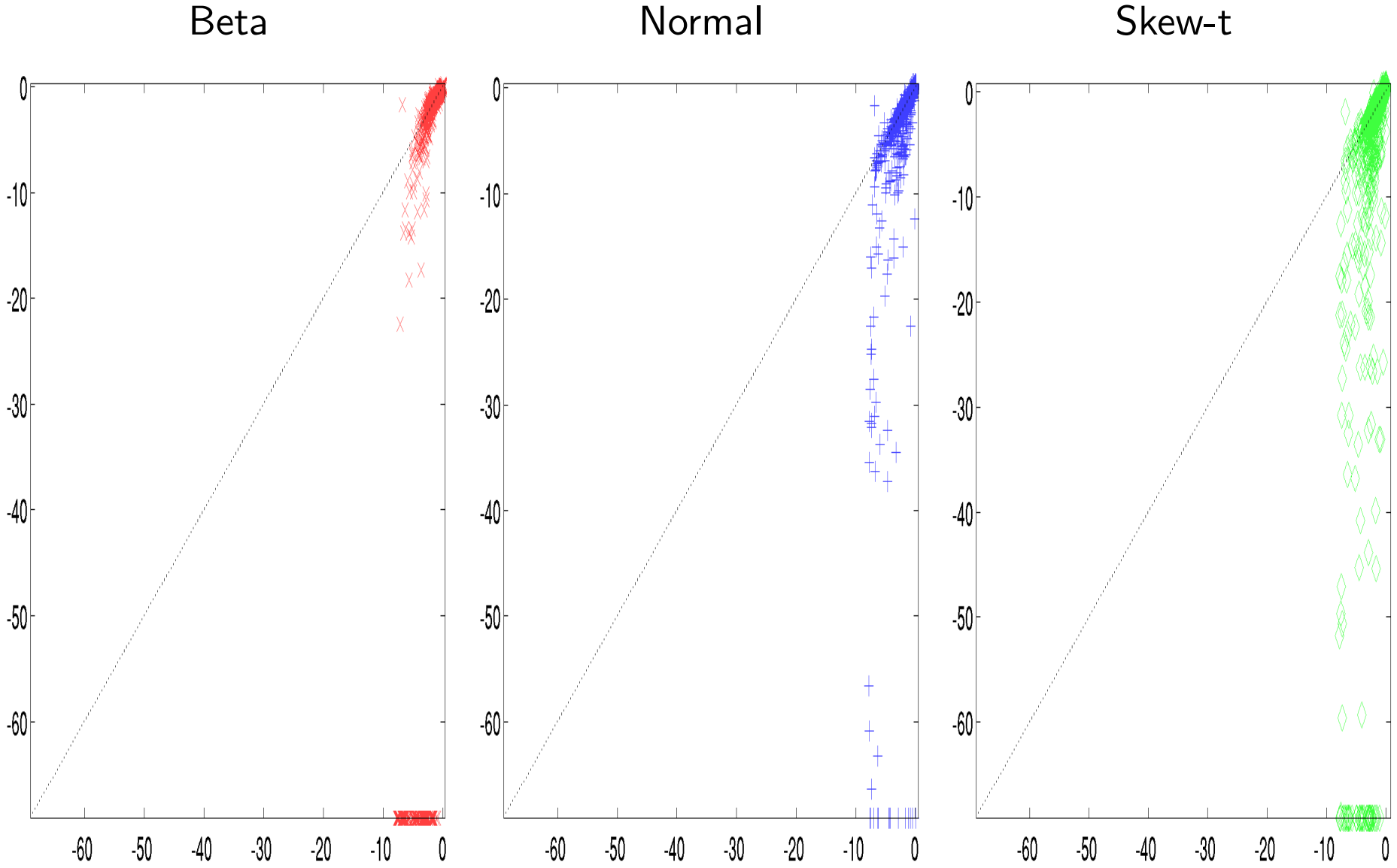
Individual Variances: Our Approach (x-axis) vs ... (y-axis)

Output Growth (H2Q2)



Individual Log Predictive Densities

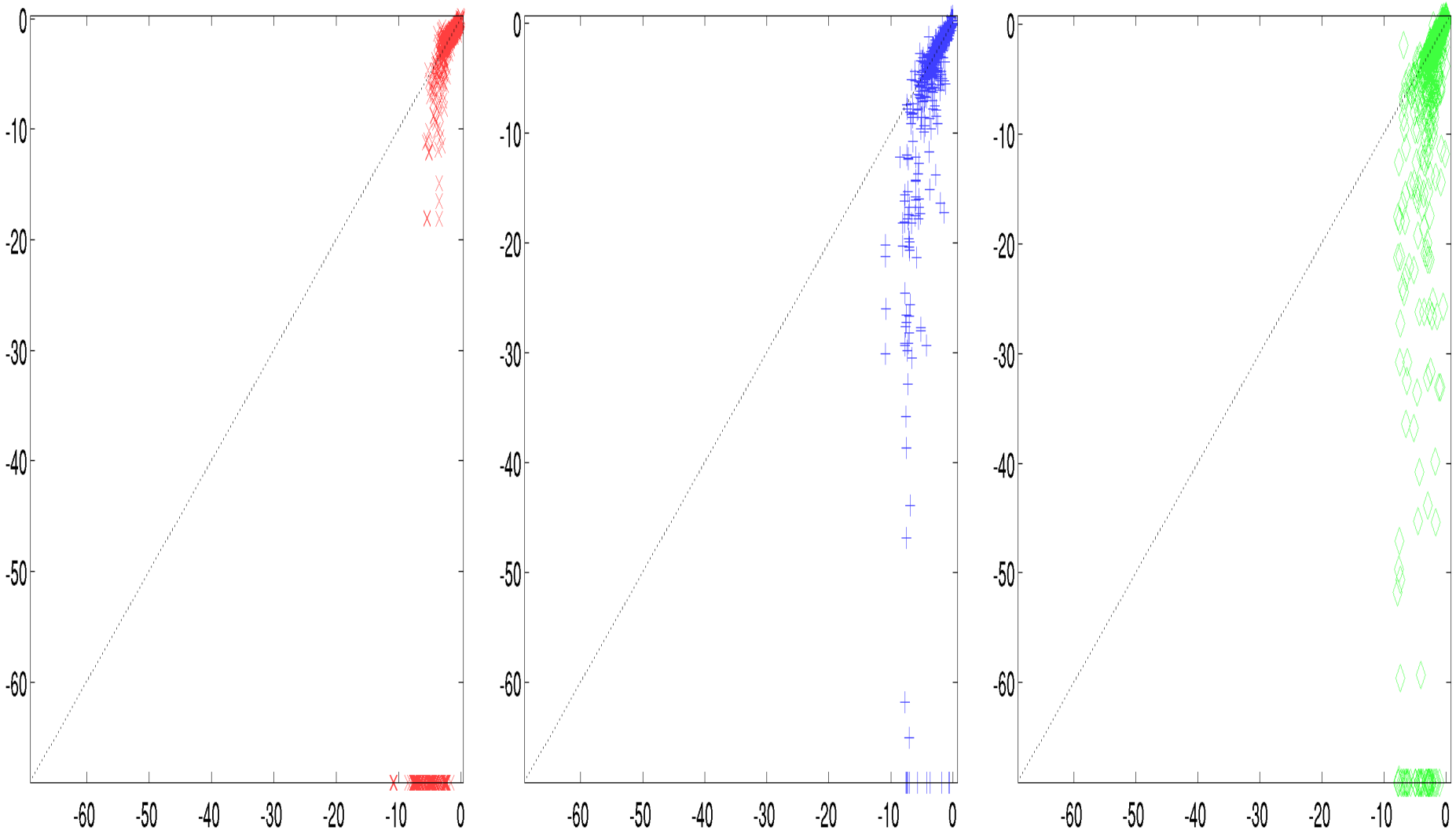
Inflation (H2Q2)



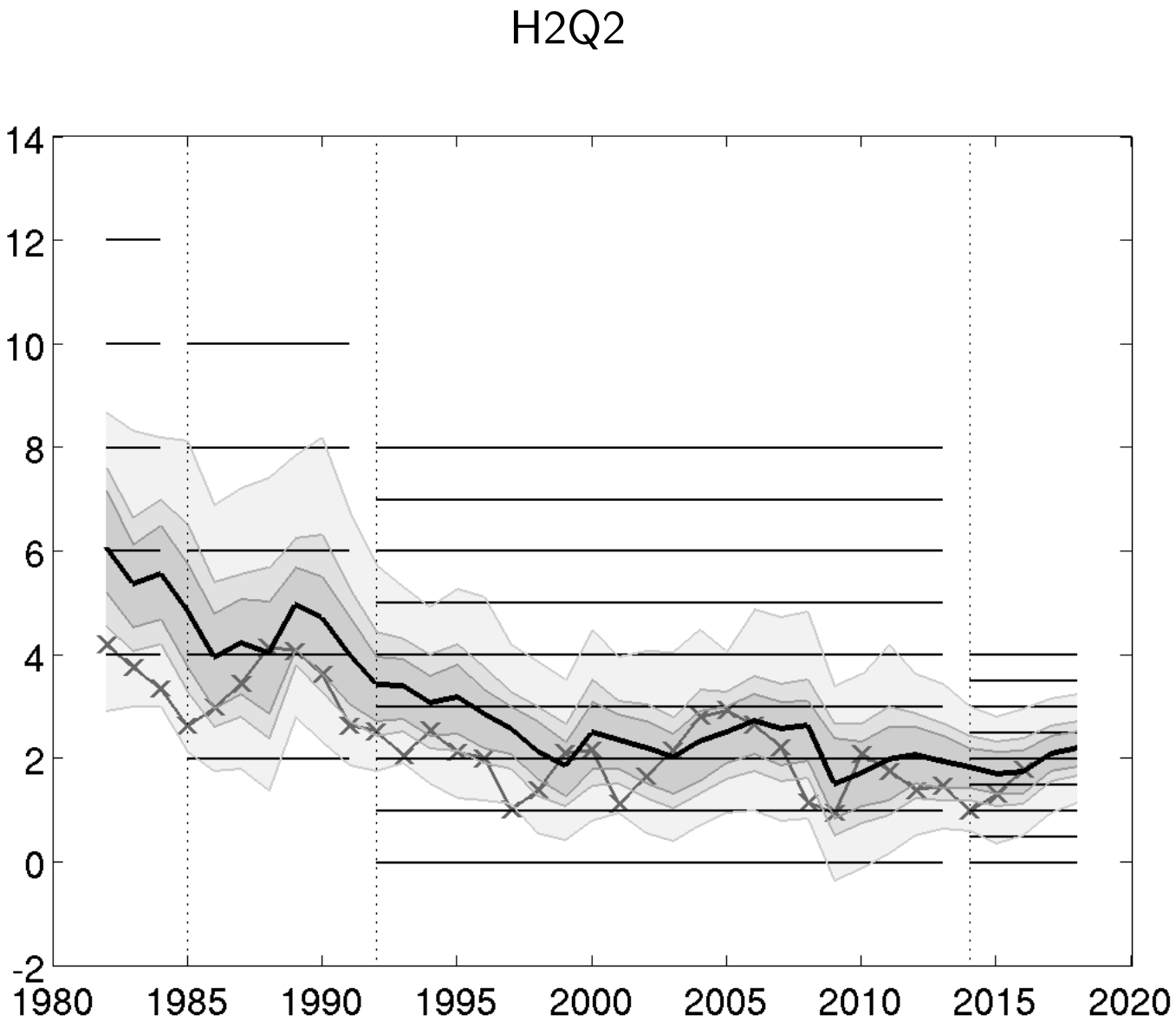
Individual Log Predictive Densities

Output Growth (H2Q2)

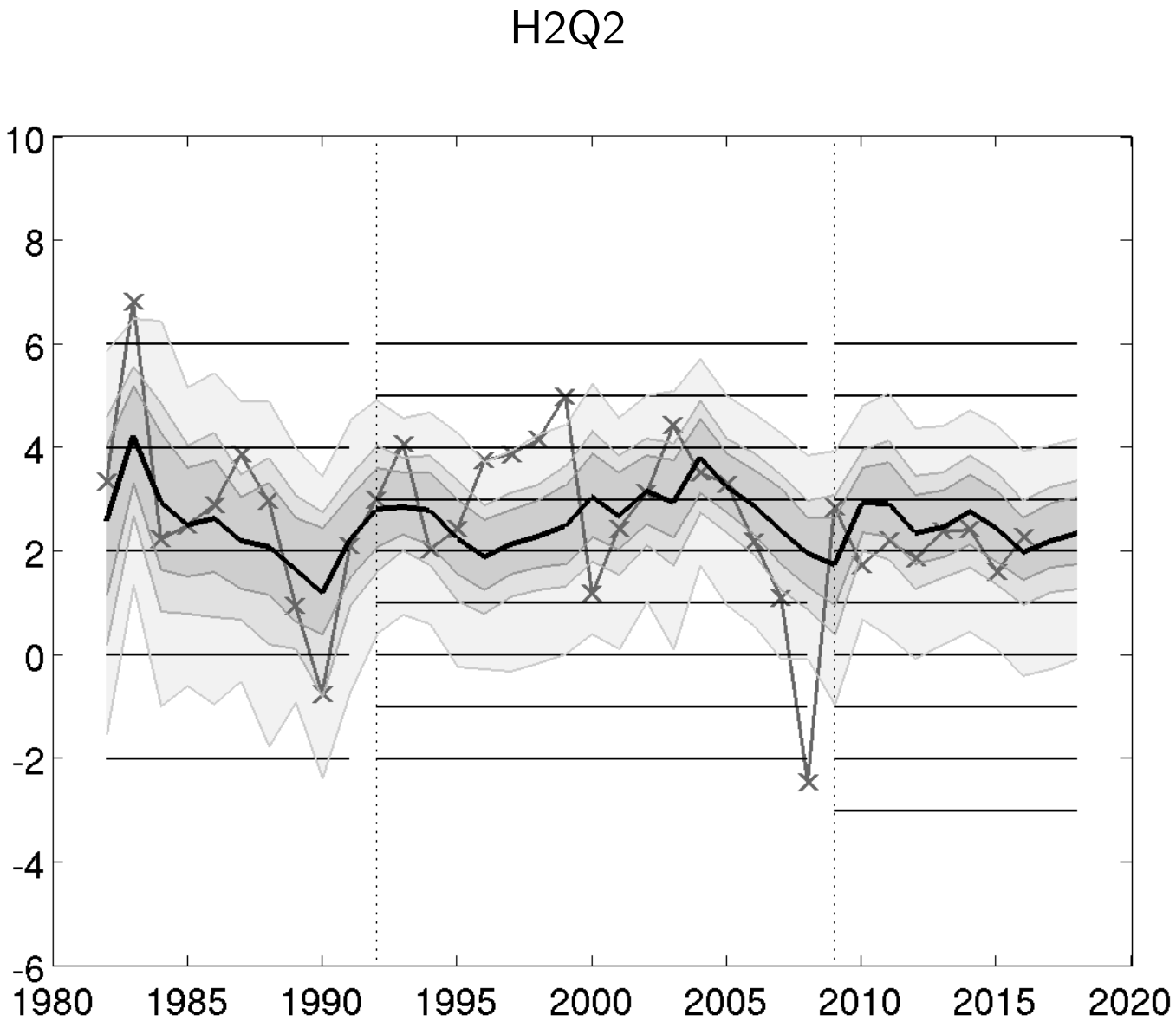
Beta | Normal | Skew-t



Average Forecast Density for Inflation Over Time



Average Forecast Density for Output Growth Over Time



The Conquest of Inflation Credibility in the U.S.

- A riff on Sargent's "The Conquest of American Inflation". That seminal book describes how inflation rose and then fell (was conquered).
- Our analysis of SPF focuses on the **evolution of forecasters' uncertainty about inflation**, and shows that:
  - The variance of the aggregate distribution has declined substantially
  - This decline mainly reflects a fall in the average variance across forecasters, as opposed to less disagreement

The Conquest of Inflation Credibility in the U.S.

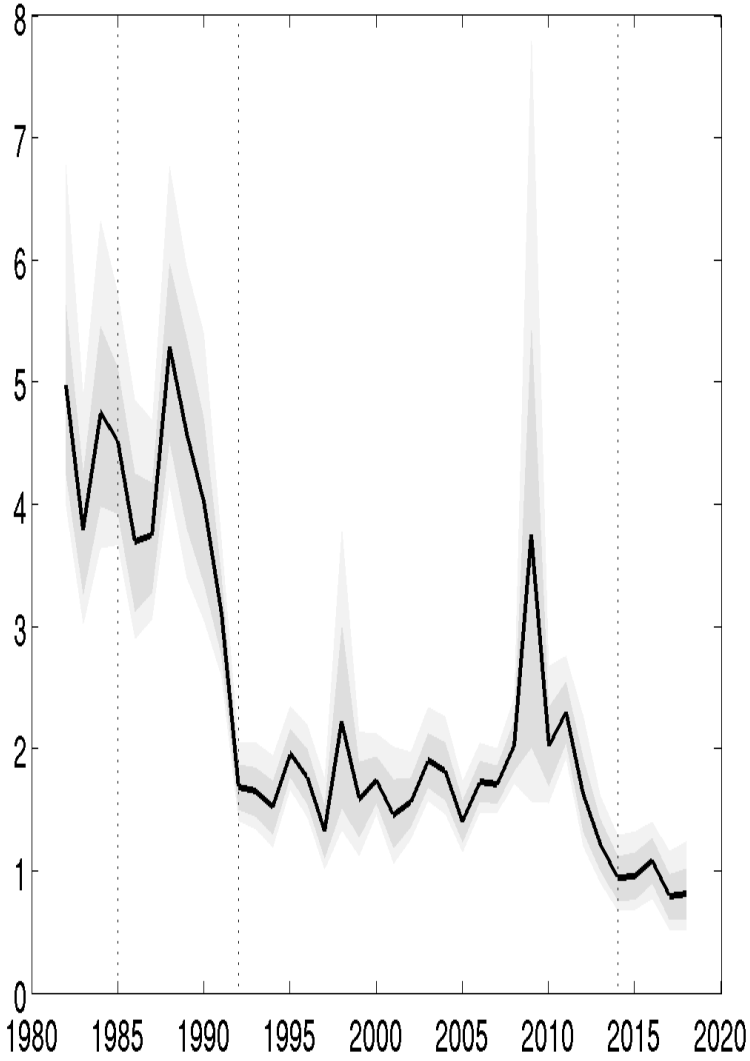
- A riff on Sargent's "The Conquest of American Inflation". That seminal book describes how inflation rose and then fell (was conquered).
- Our analysis of SPF focuses on the **evolution of forecasters' uncertainty about inflation**, and shows that:
  - The variance of the aggregate distribution has declined substantially
  - This decline mainly reflects a fall in the average variance across forecasters, as opposed to less disagreement
  - There is strong suggestive evidence that the decline is associated with the changes in the bins
  - In other words, forecasters views on the extent to which the central bank has inflation under control may have not changed much since the 1980s, only the way in which the Philadelphia Fed has been asking the question has changed

The Conquest of Inflation Credibility in the U.S.

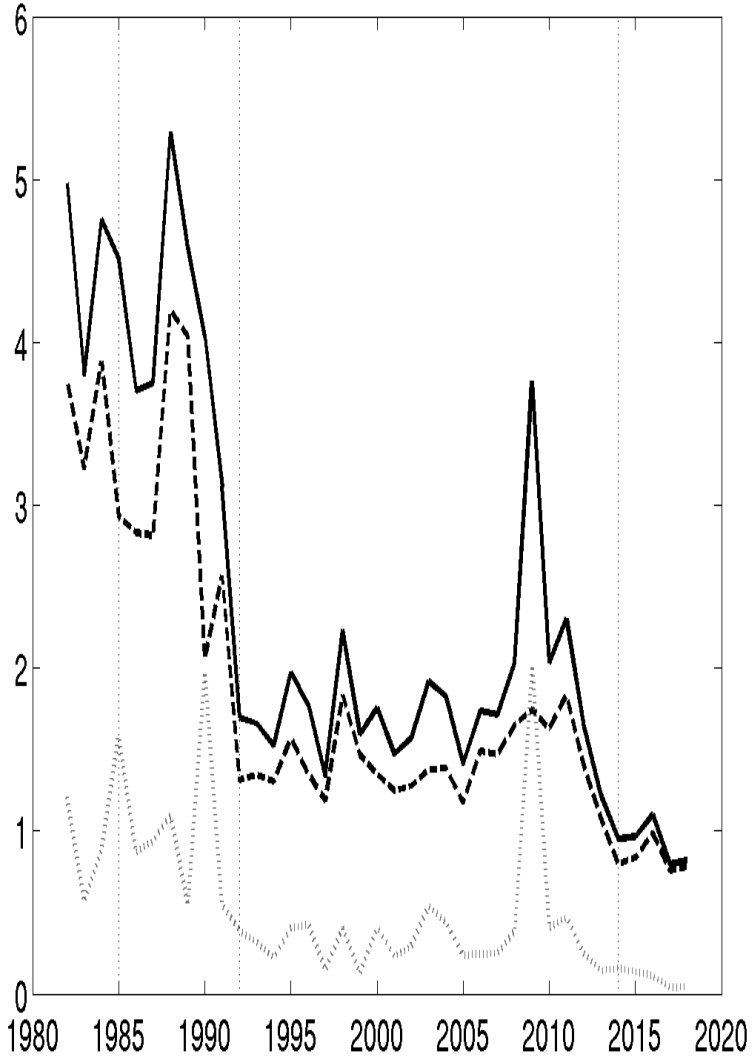
- A riff on Sargent's "The Conquest of American Inflation". That seminal book describes how inflation rose and then fell (was conquered).
- Our analysis of SPF focuses on the **evolution of forecasters' uncertainty about inflation**, and shows that:
  - The variance of the aggregate distribution has declined substantially
  - This decline mainly reflects a fall in the average variance across forecasters, as opposed to less disagreement
  - There is strong suggestive evidence that the decline is associated with the changes in the bins
  - In other words, forecasters views on the extent to which the central bank has inflation under control may have not changed much since the 1980s, only the way in which the Philadelphia Fed has been asking the question has changed
  - At the same time we show that the likelihood of extreme events (e.g., "back to the seventies," or deflation) has declined over time

The "Conquest of Inflation Credibility in the U.S."

Variance

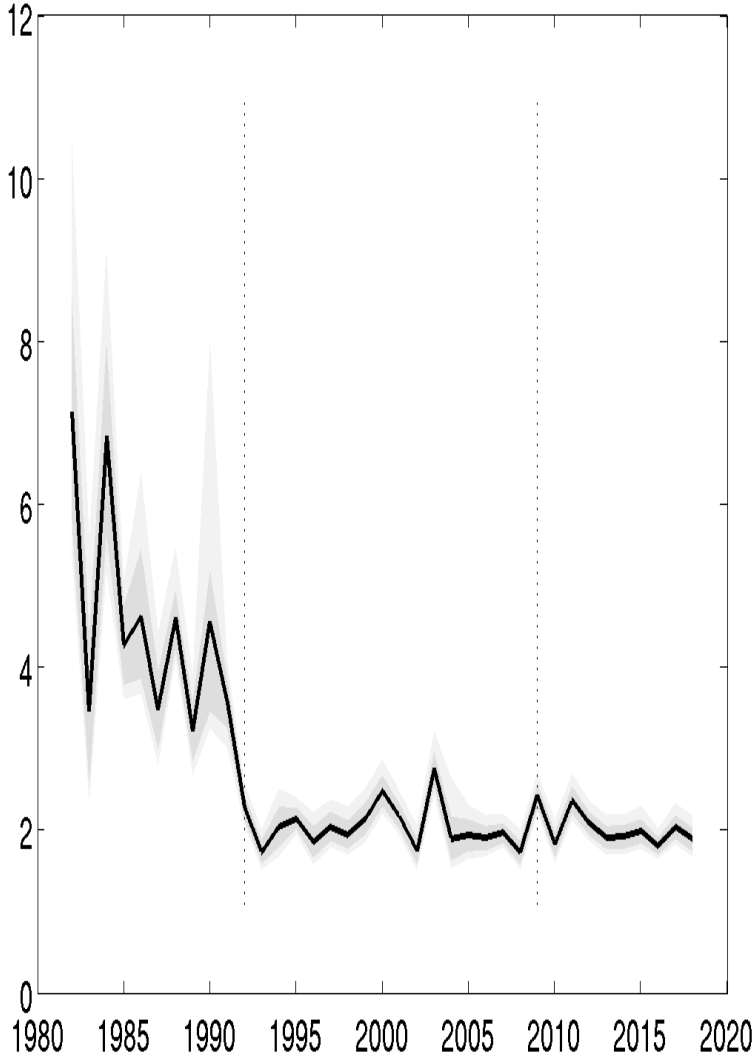


Uncertainty and Disagreement

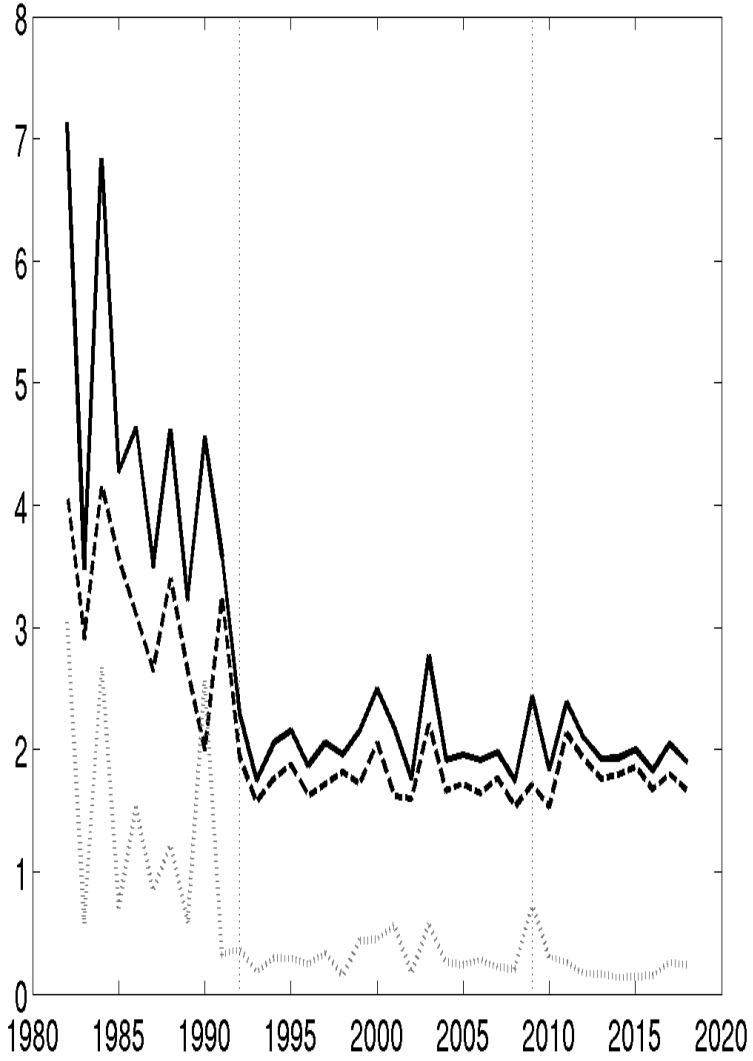


Variance, Uncertainty, and Disagreement for Output

Variance

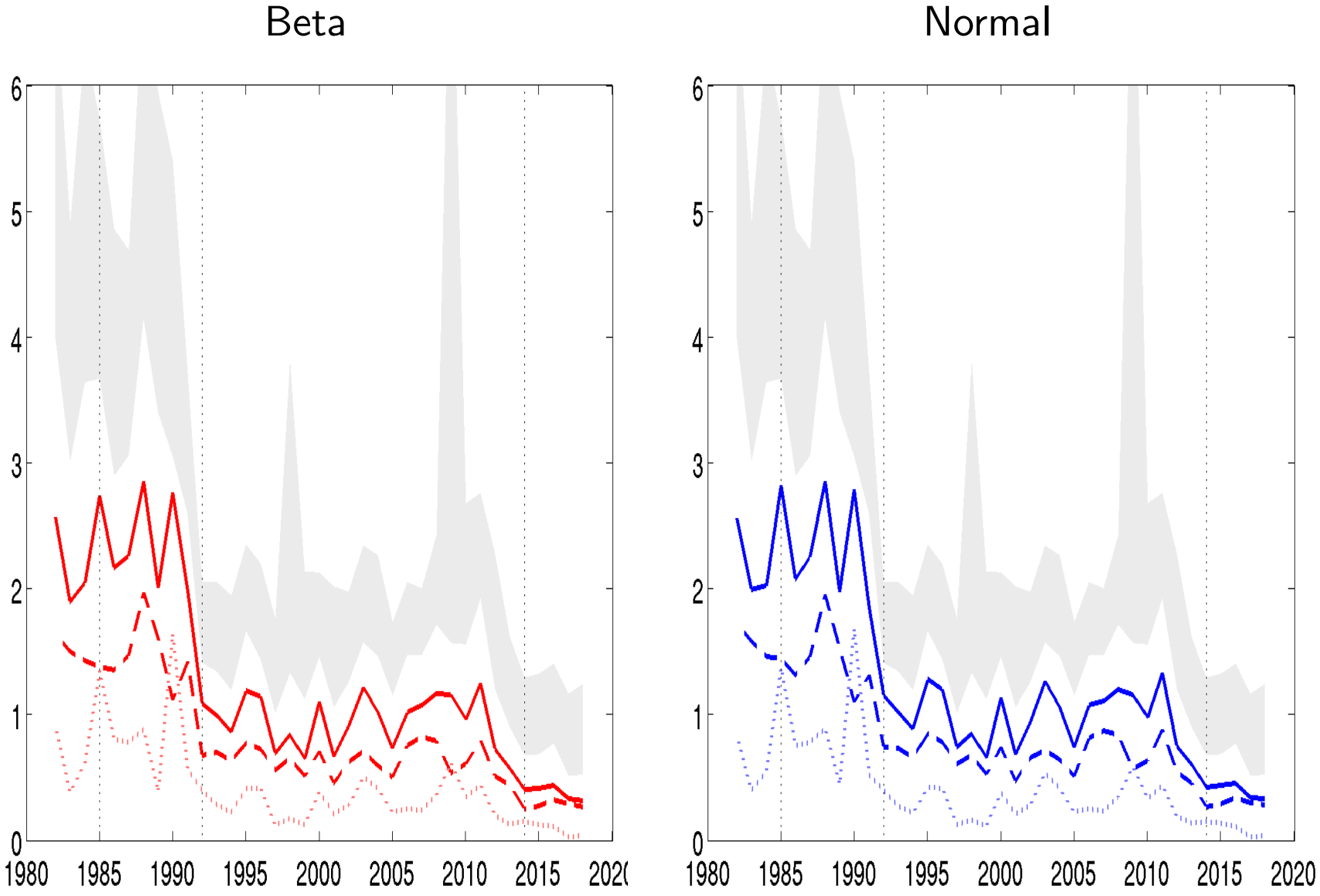


Uncertainty and Disagreement



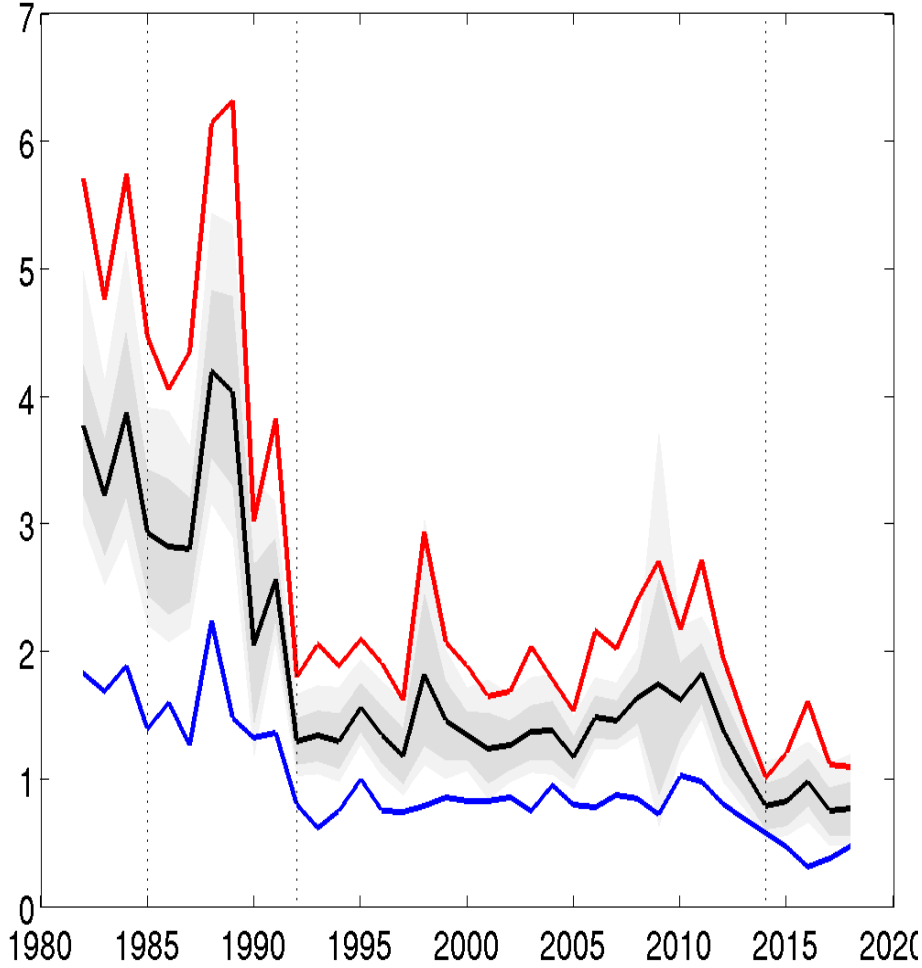
Variance, Uncertainty, and Disagreement: Alt. Methods

Inflation (H2Q2)

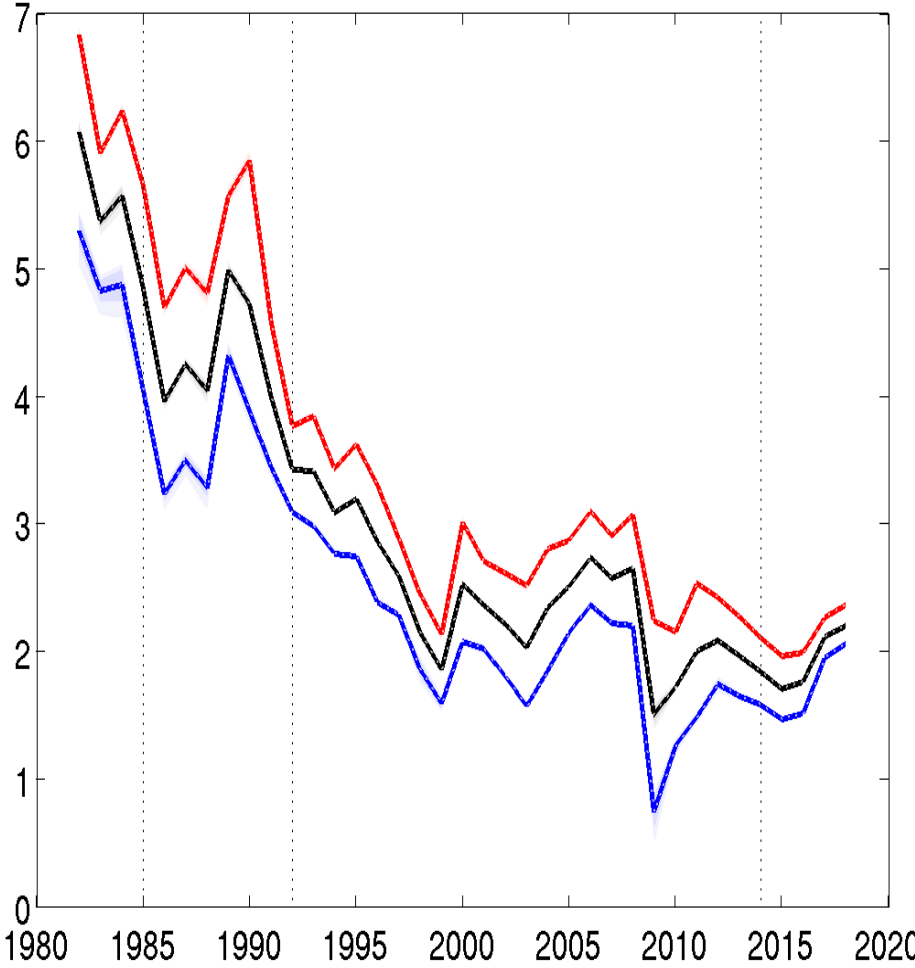


Heterogeneity in Means and Variances

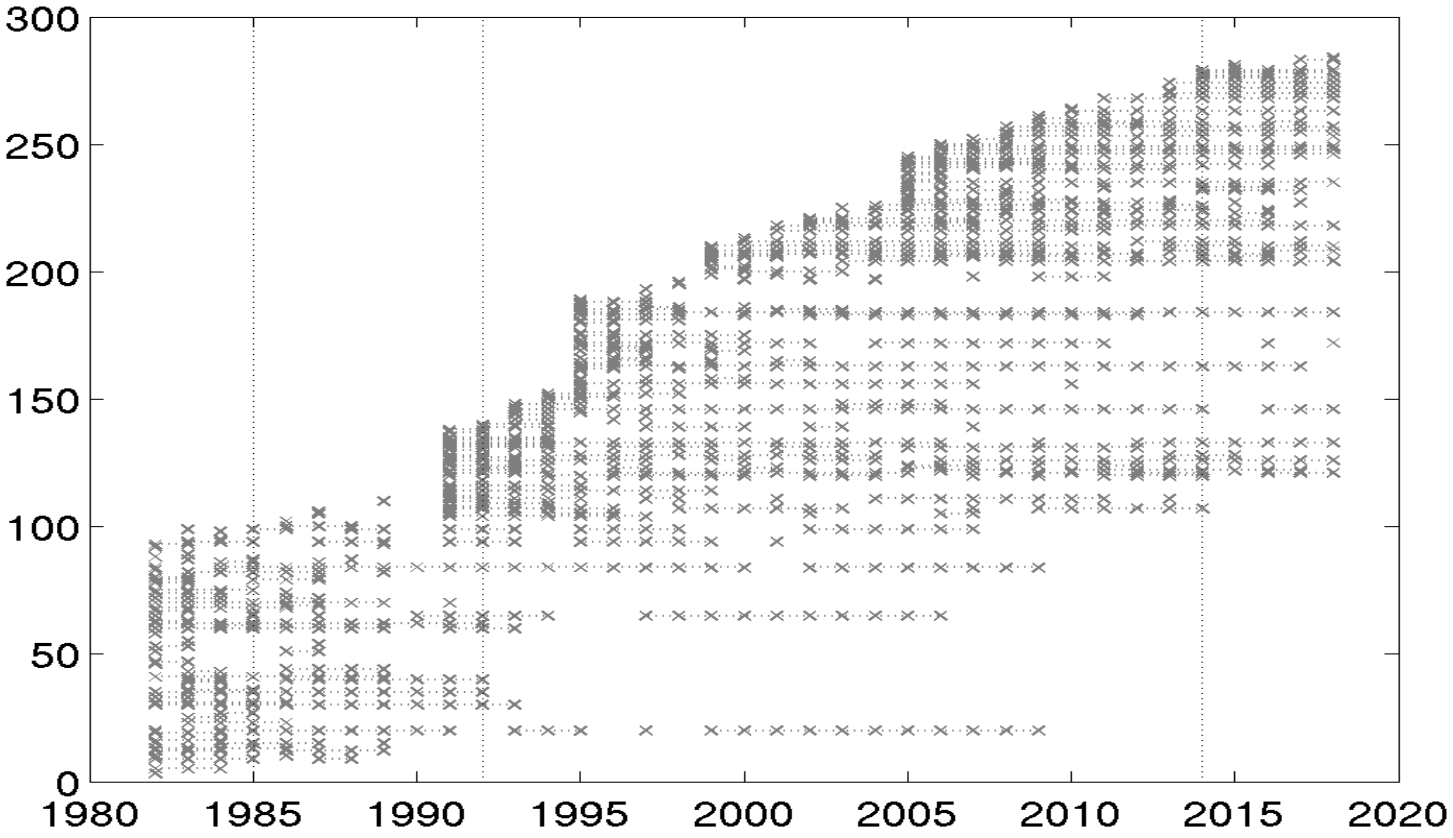
Variance – High vs Low



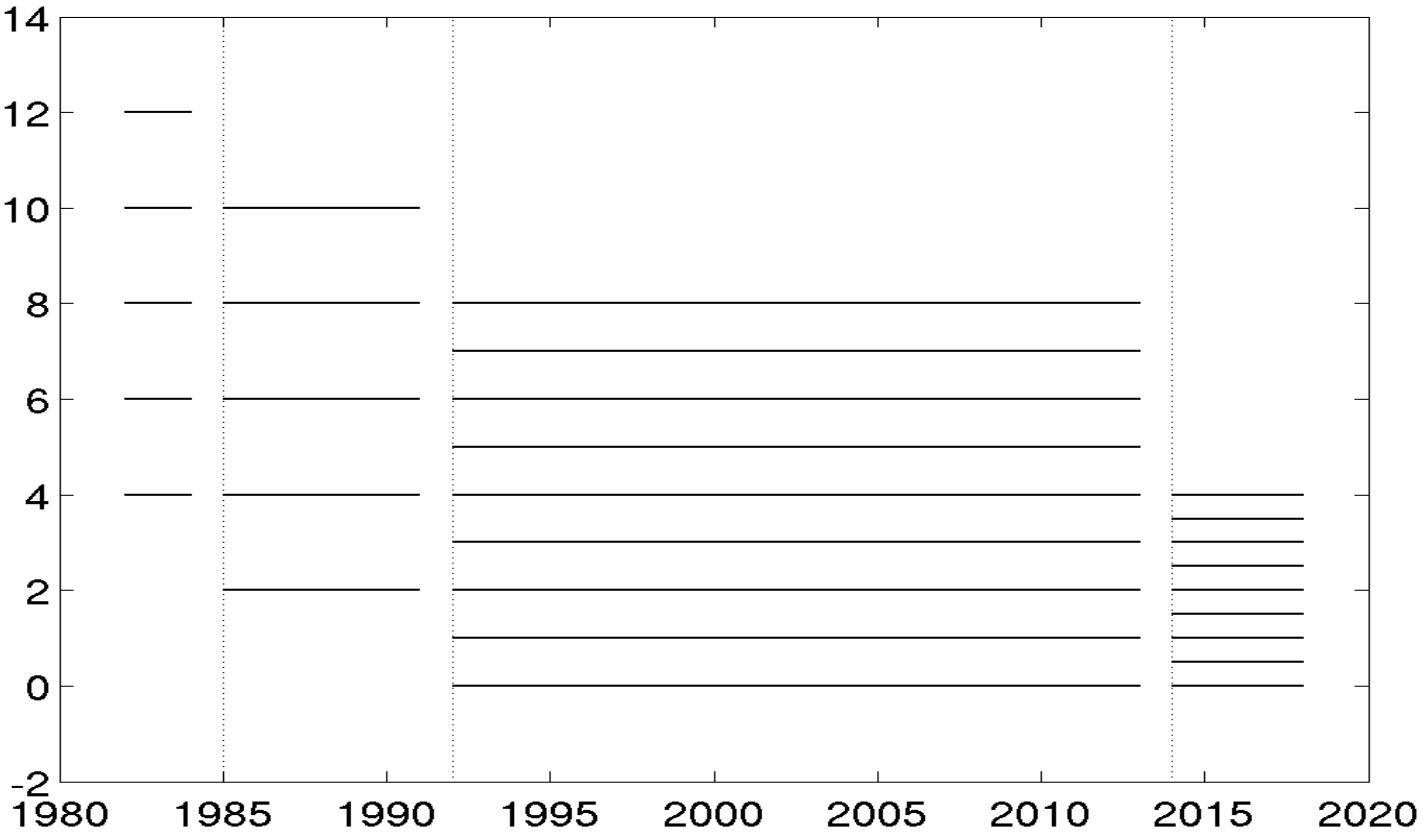
Mean – High vs Low



Forecasters

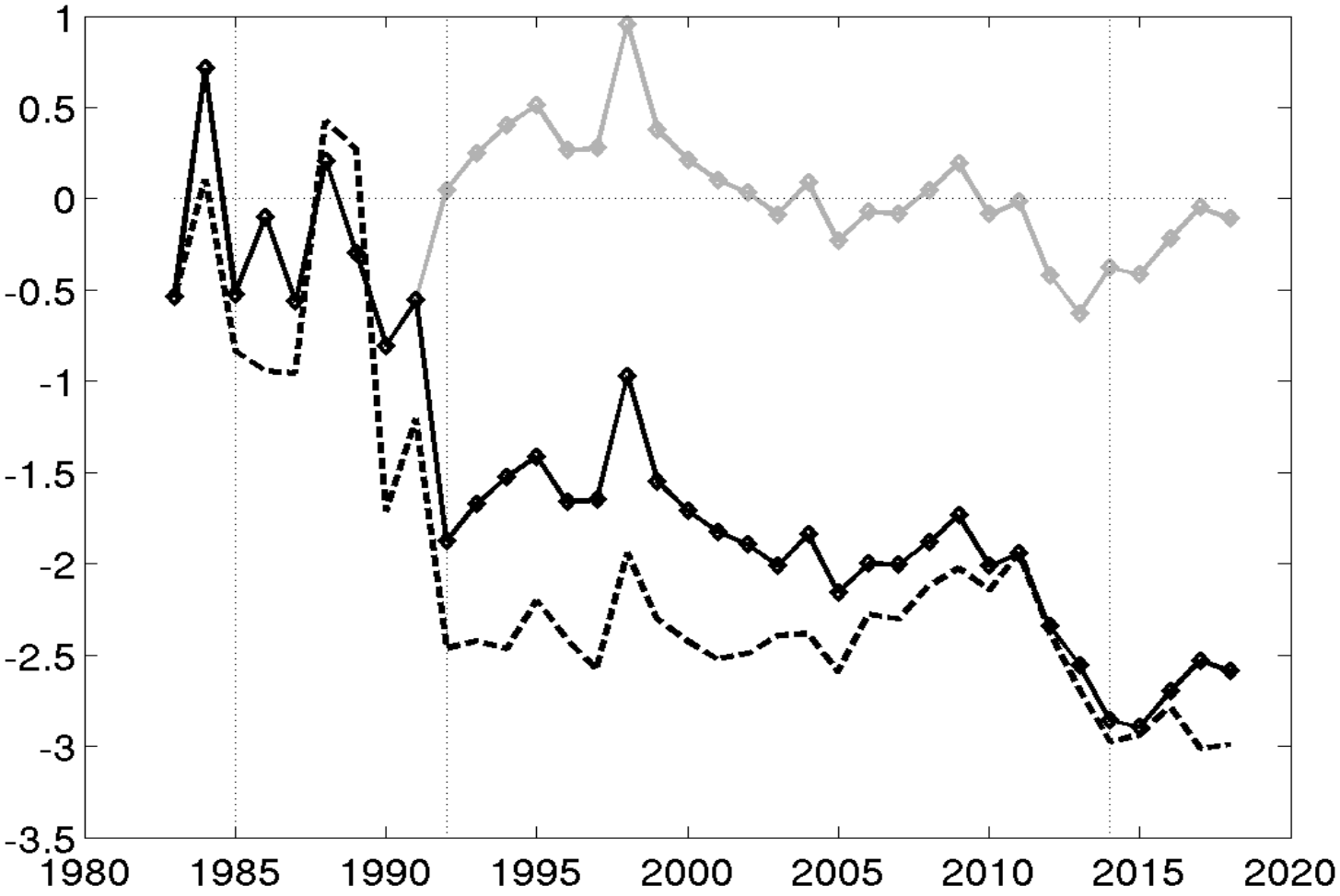


Bins



The Role of Survey Composition and Design

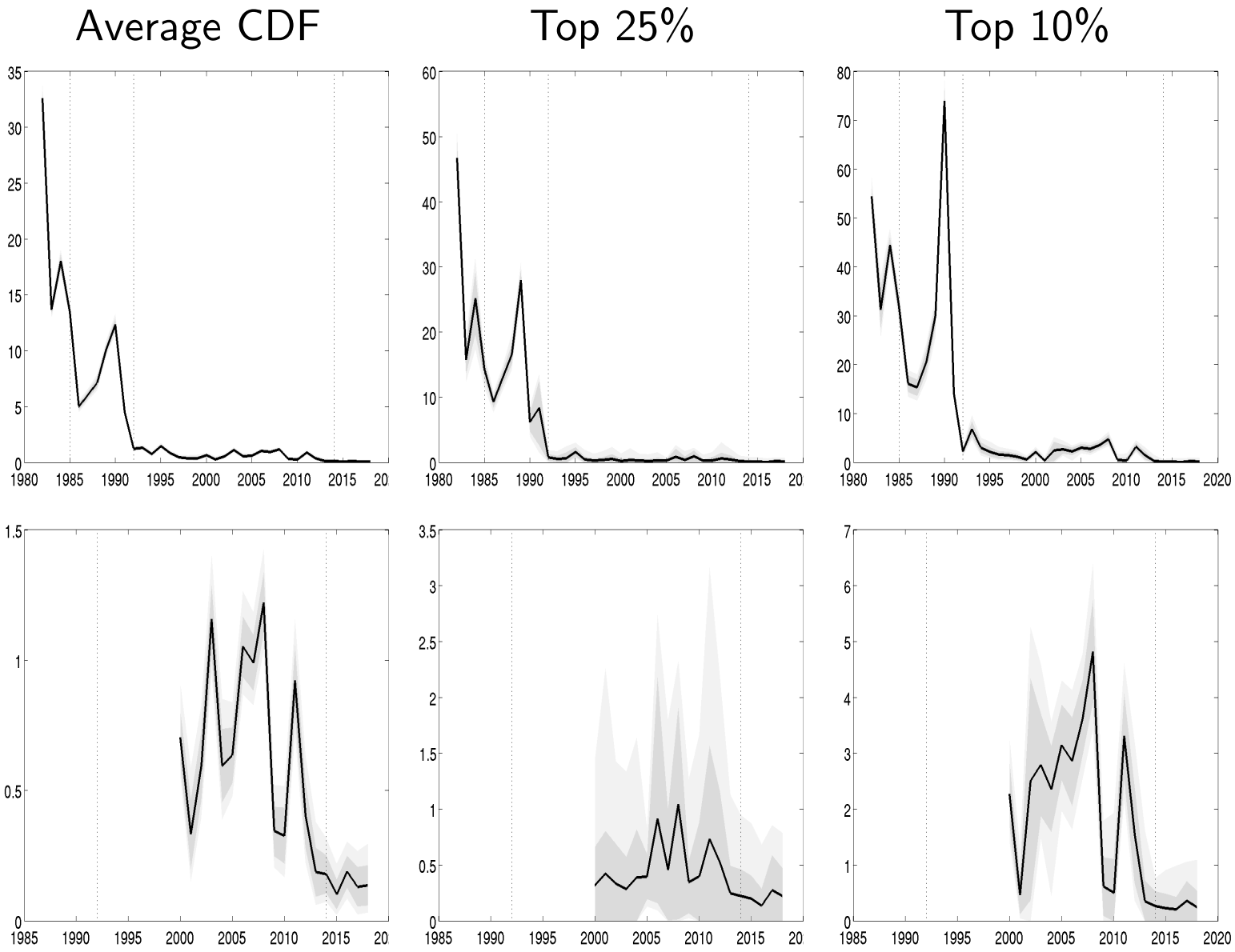
Cumulated Changes in Variance



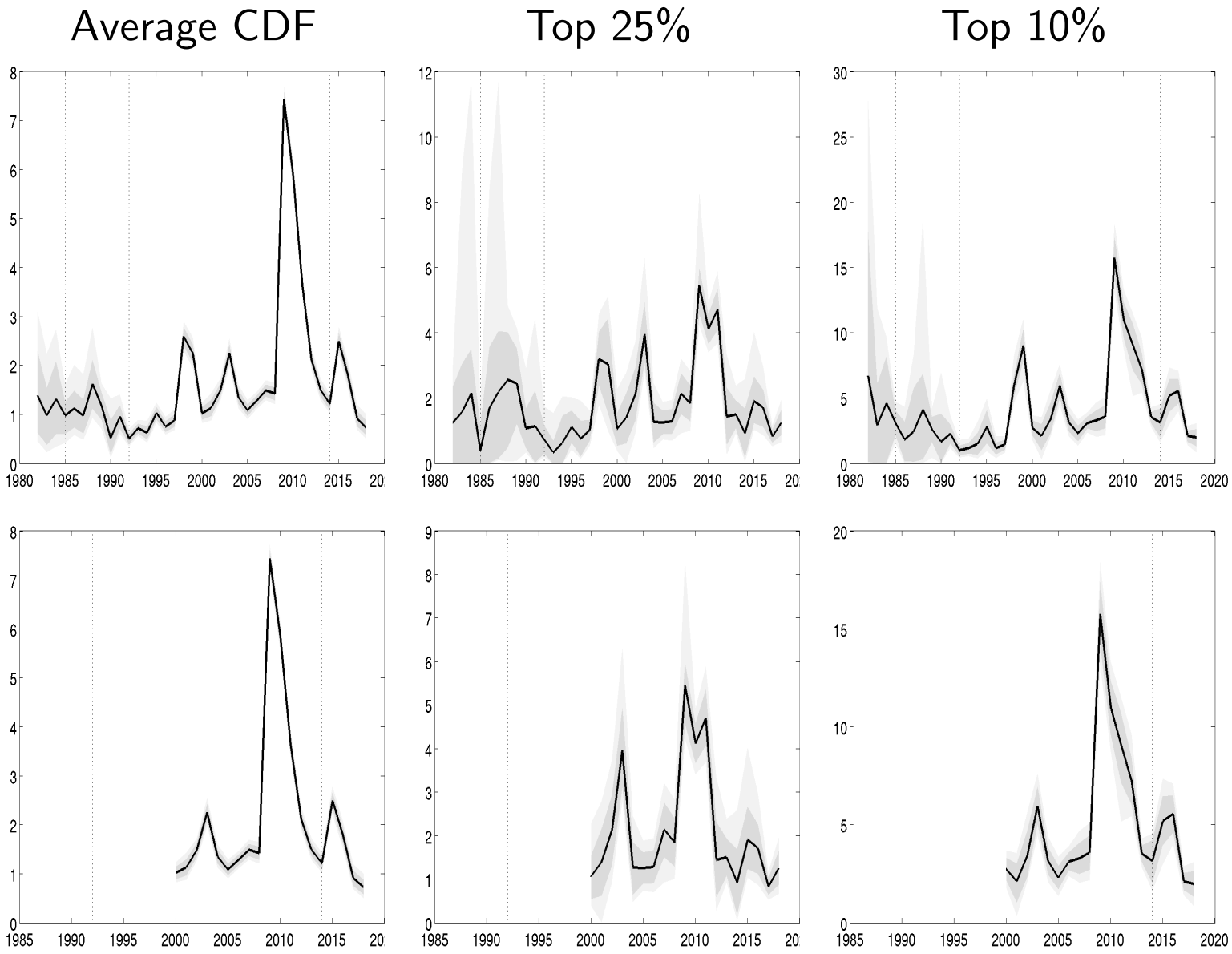
The Likelihood of Extreme Events

- What is the likelihood of "going back to the seventies" (inflation at or above 6.9%, the average for the 70s)?
    - Really small. Negligible for most forecasters. Survey design (bins composition) does not seem to drive this finding
    - No striking support in favor of DSGE models with regime switching where one regime is a "return to the seventies"
    - Caveat: only 2 years ahead forecast distribution
- Likelihood of deflation non negligible after the great recession, but small currently
- Andrade et al., 2014, perform a similar exercise using extreme quintiles of the predictive distribution ("inflation @ risk")

Extreme Events: Back to the Seventies?



Extreme Events: Deflation Fear

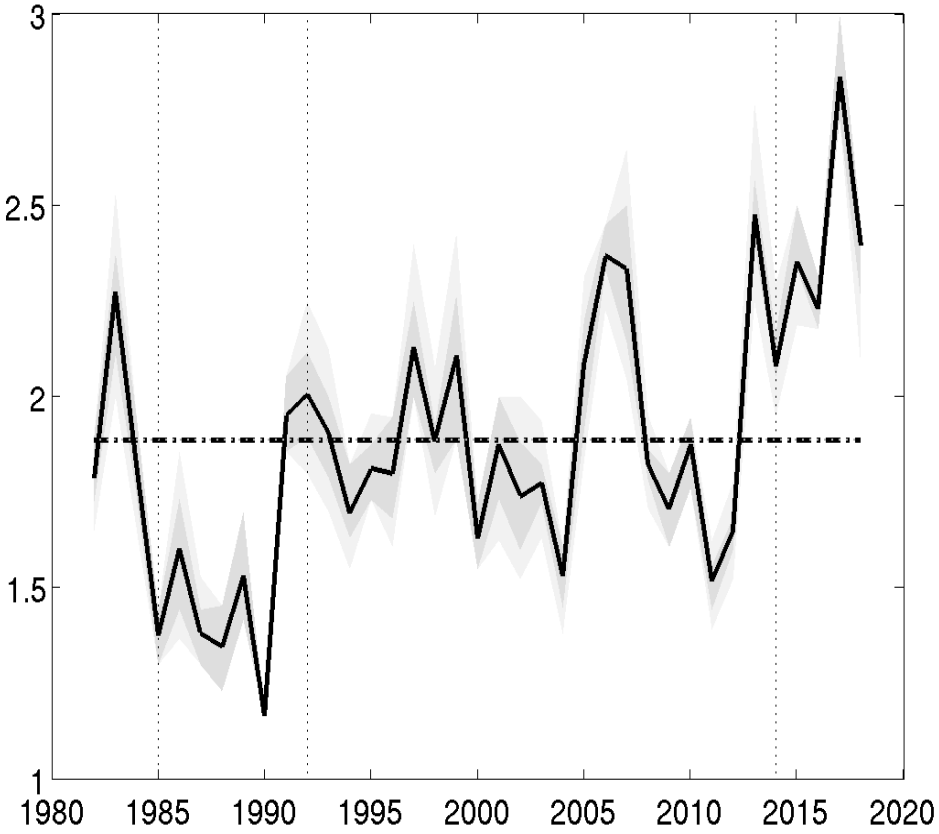


Conclusions

- We use a Bayesian Non Parametric approach to conduct inference on the subjective probability distributions for two-year ahead inflation for the Survey of Professional Forecasters.

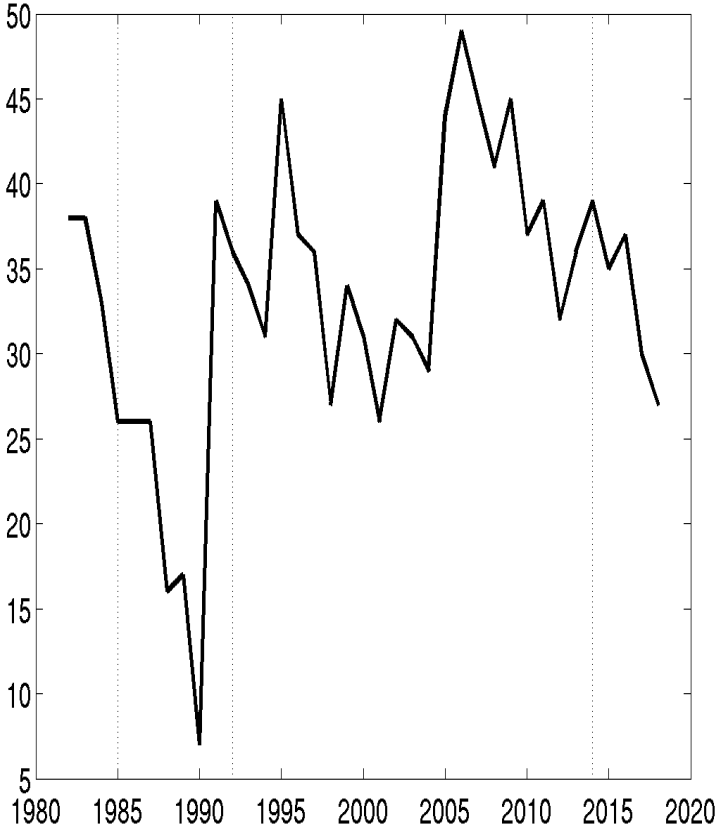
- Findings:
  - Current approaches lead to an *underestimation of the variance* of forecasters.
  - Forecasters views on the extent to which the central bank has inflation under control may have not changed much since the 1980s, only the way in which the Philadelphia Fed has been asking the question has changed
  - Extreme events: likelihood of "going back to the seventies" has been really small over the past twenty years; likelihood of deflation high after the great recession, but also small currently

Number of Forecasters/Number of Clusters

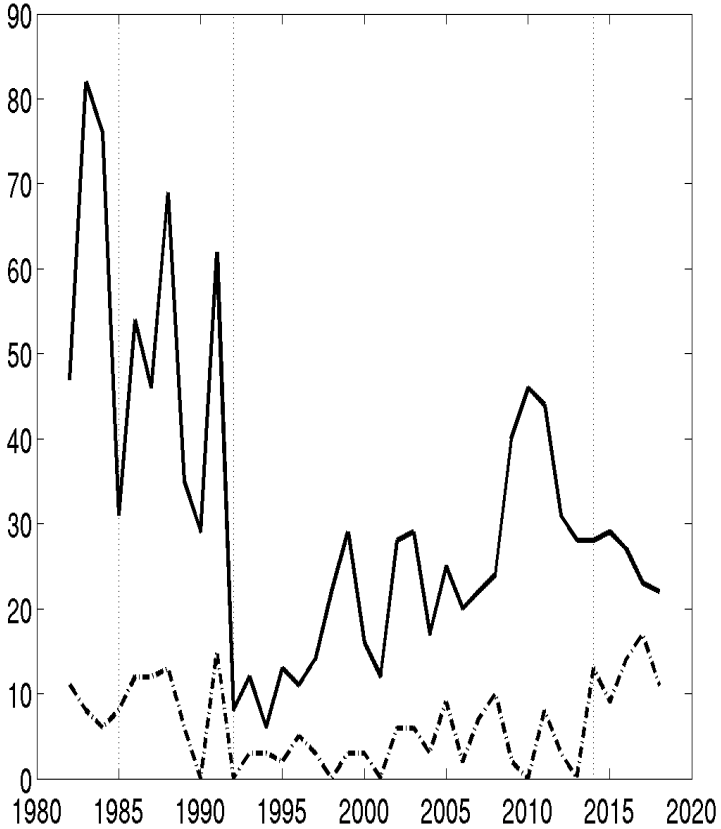


Number of Respondents and Percentage of Respondents with Probability on Open Bins (lower/higher or both)

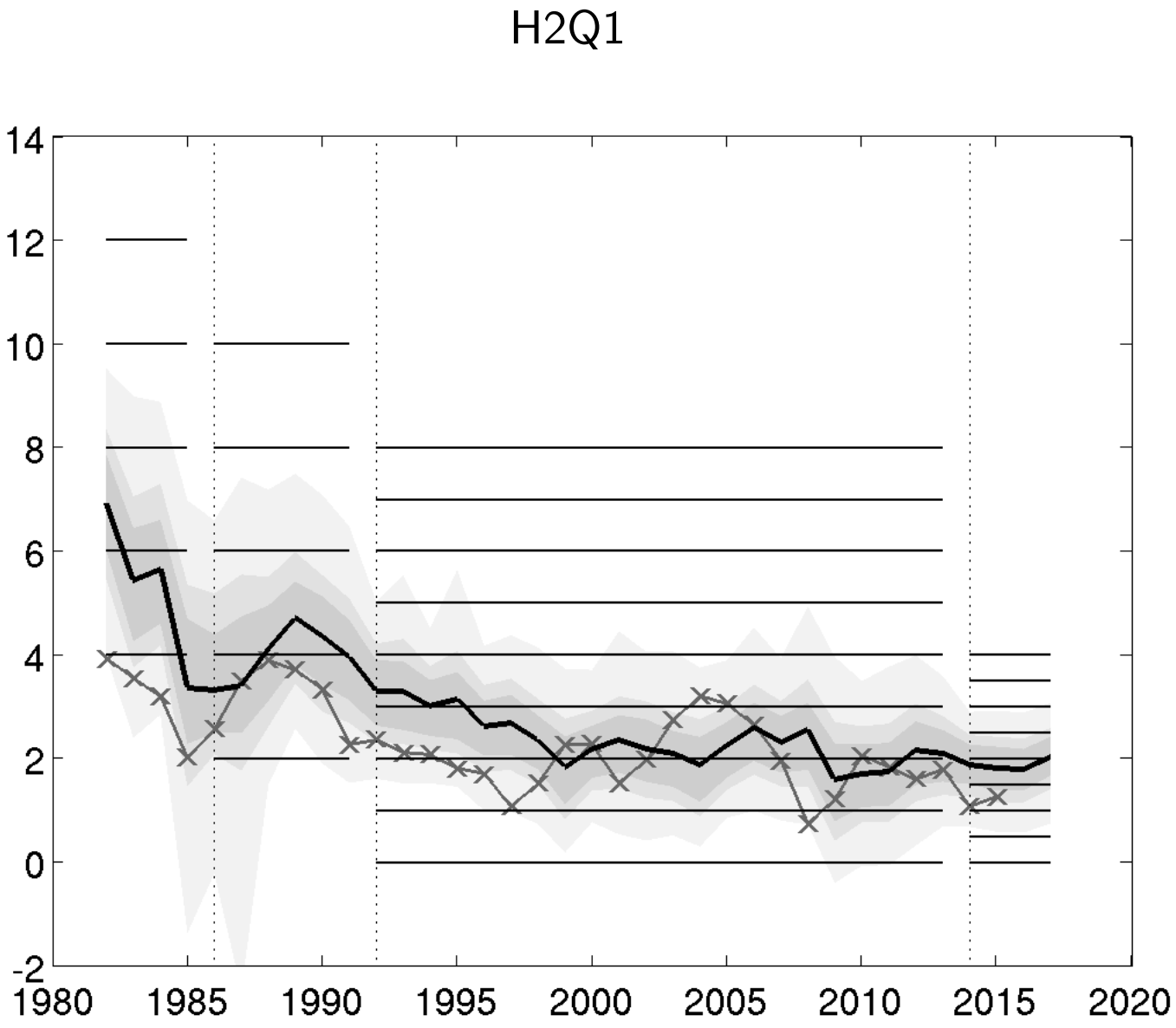
Number of Respondents



Percentage of Respondents with Probability on Open Bins

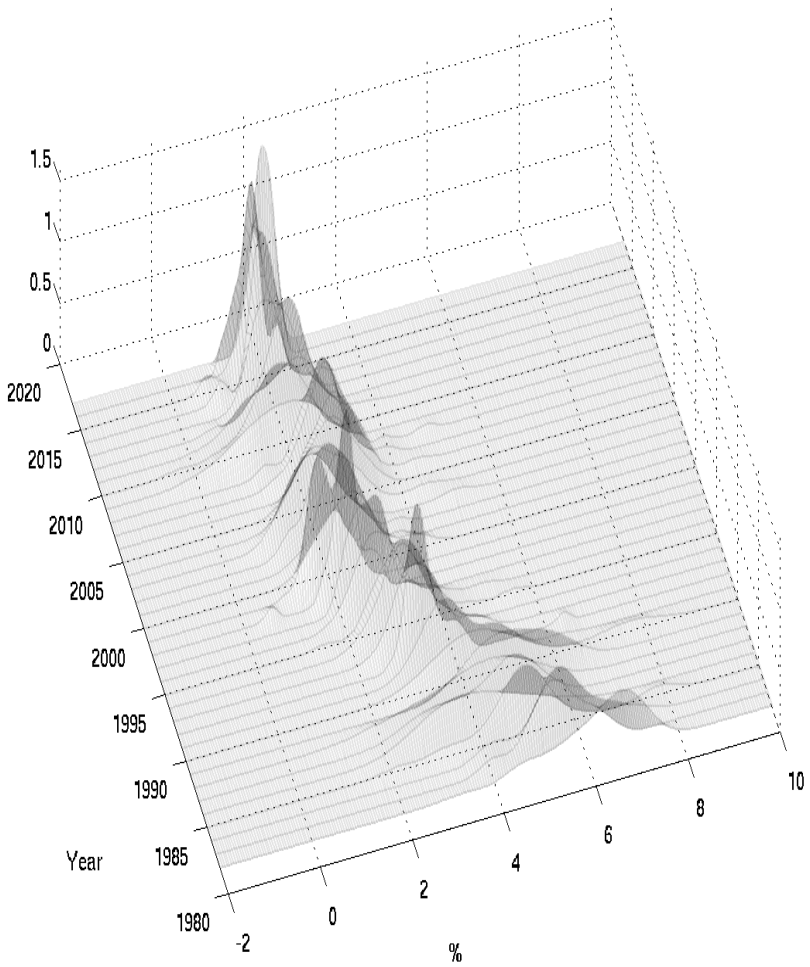


Average Forecast Density for Inflation Over Time

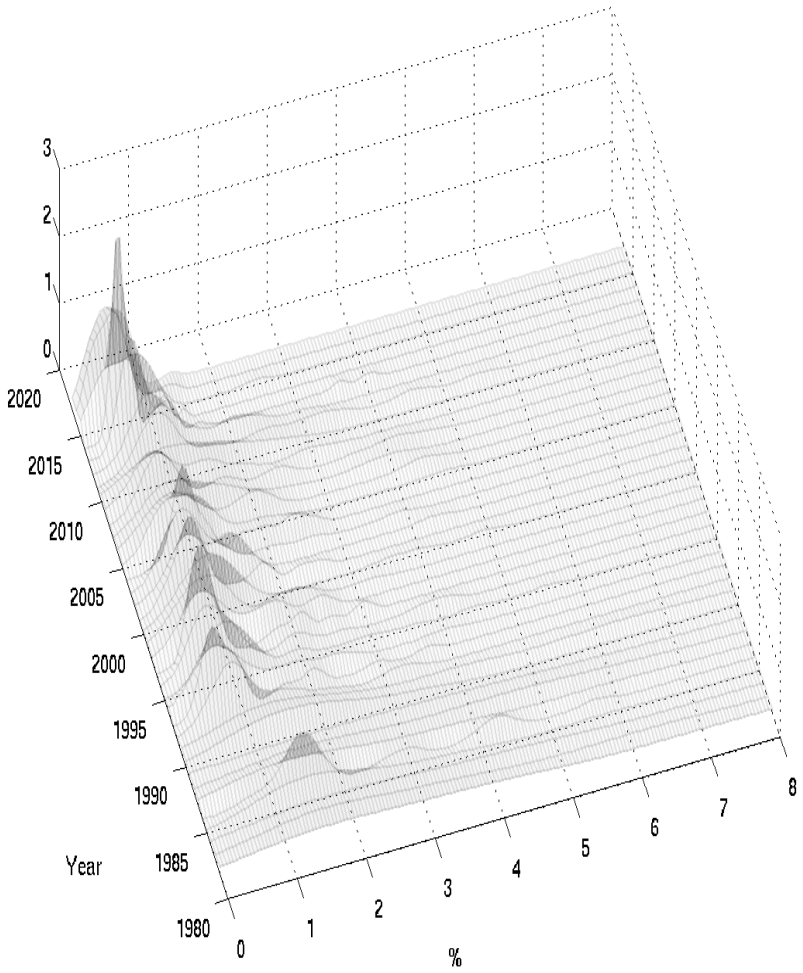


Evolution Over Time in the Cross-Sectional Distribution of Means and Variances

Mean



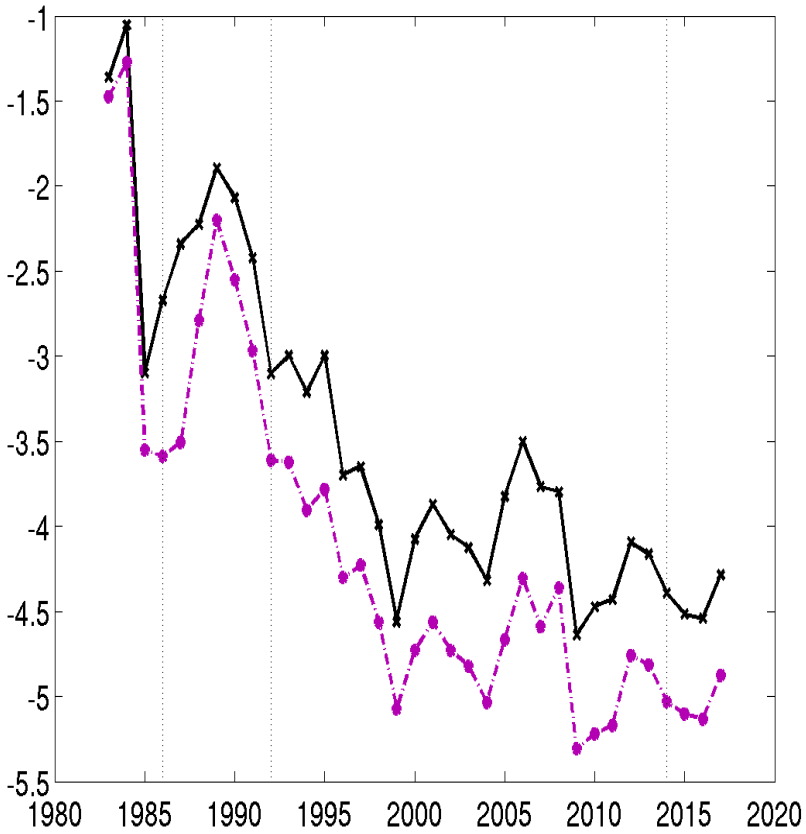
Variance



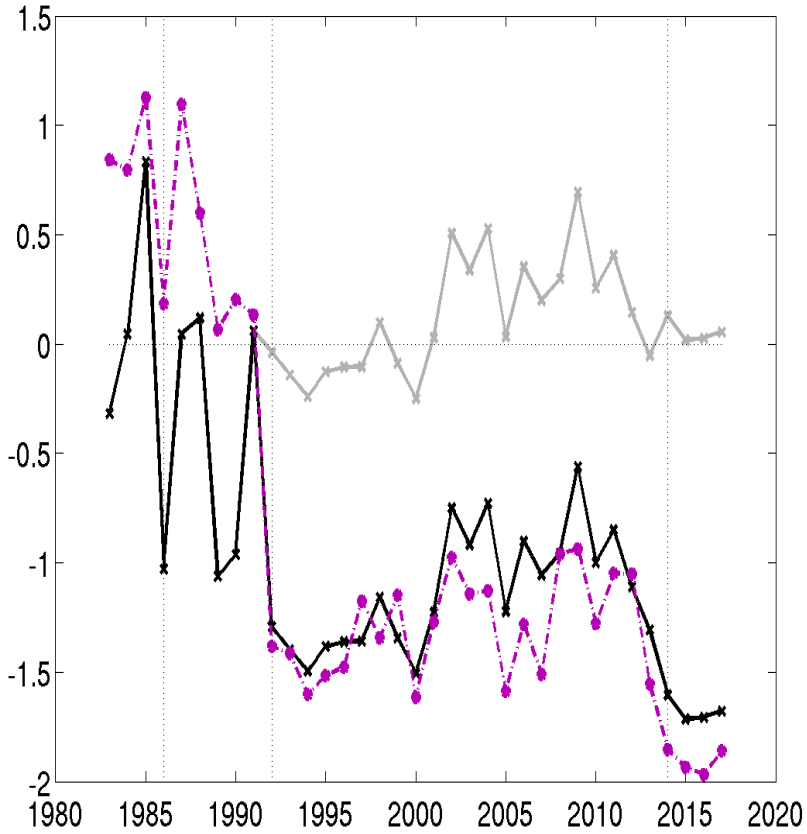
Changes in Forecasts vs Survey Composition (H2Q1)

Cumulated Changes in

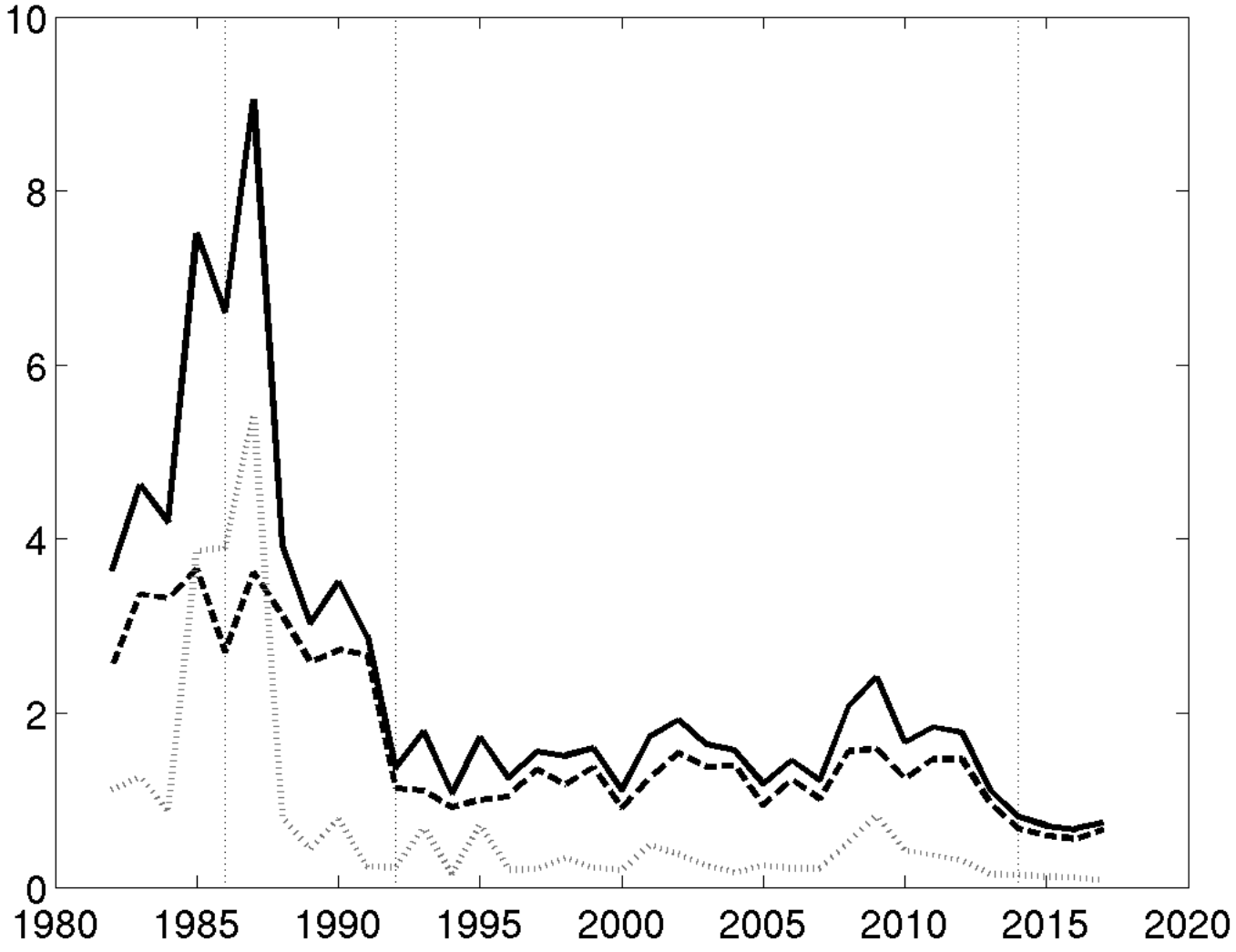
Mean



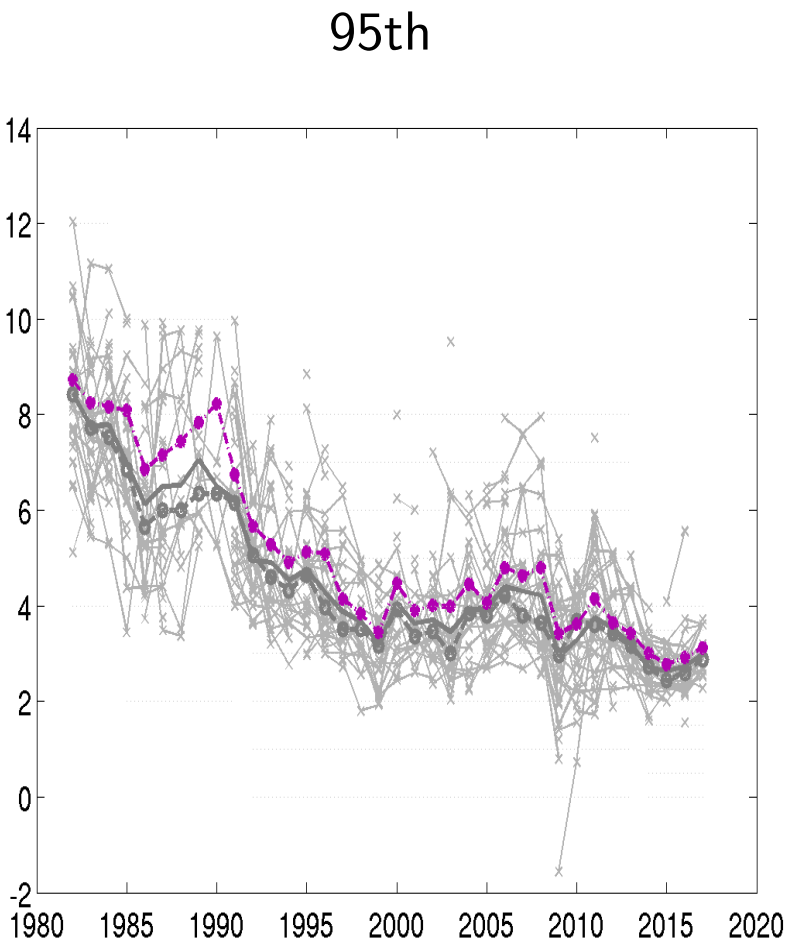
Variance

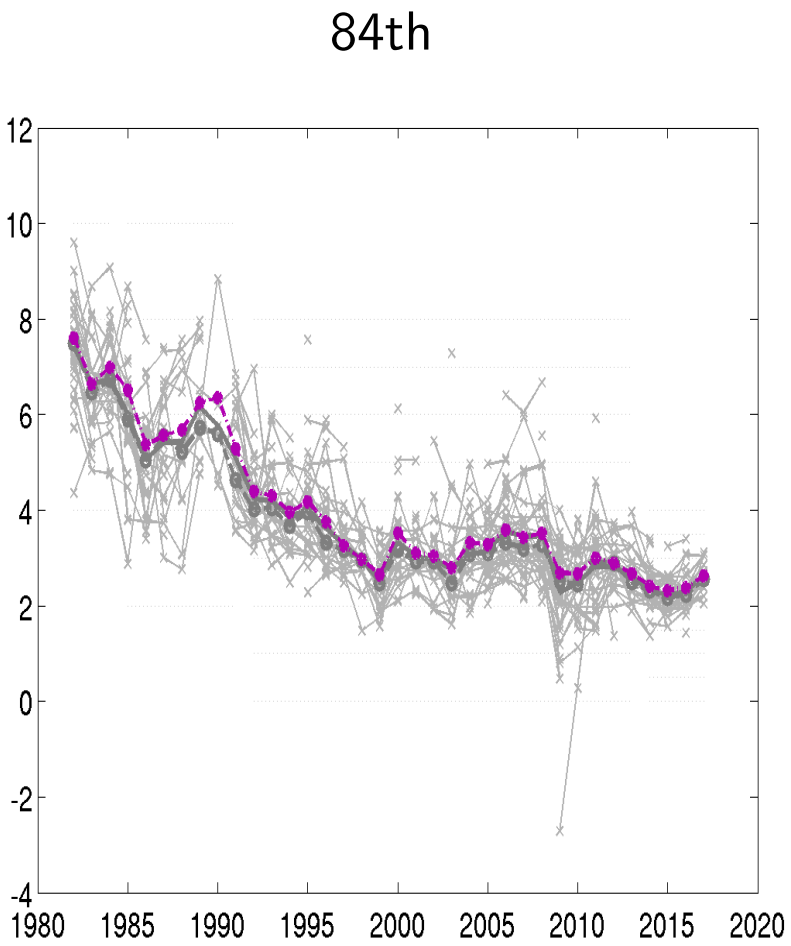


Uncertainty or Disagreement? (H2Q1)

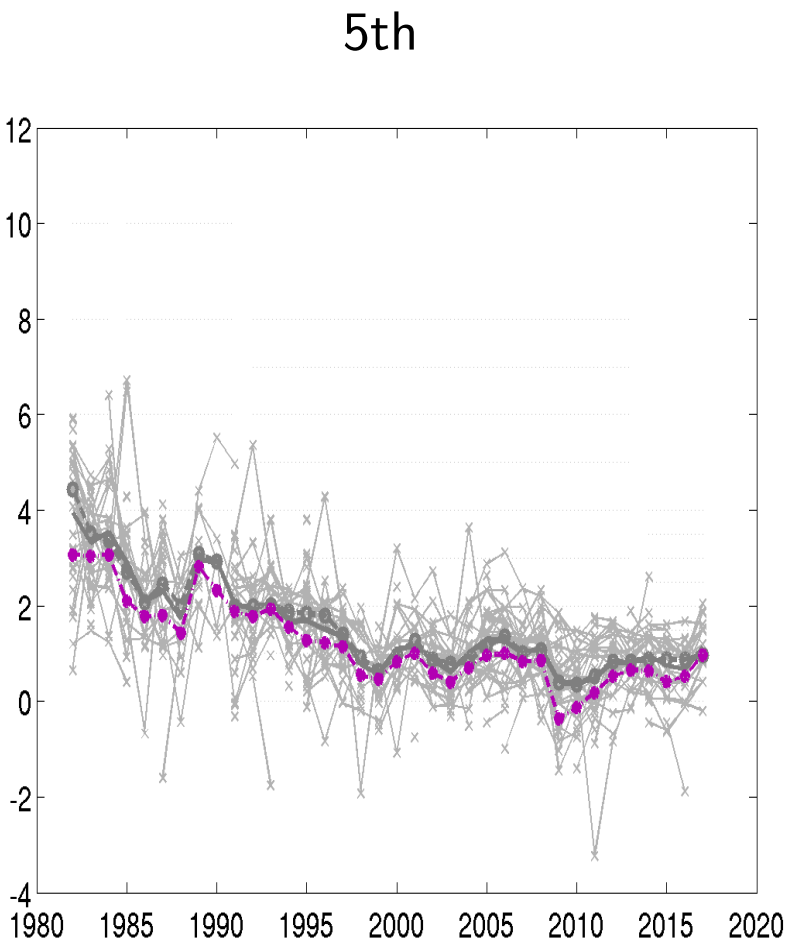


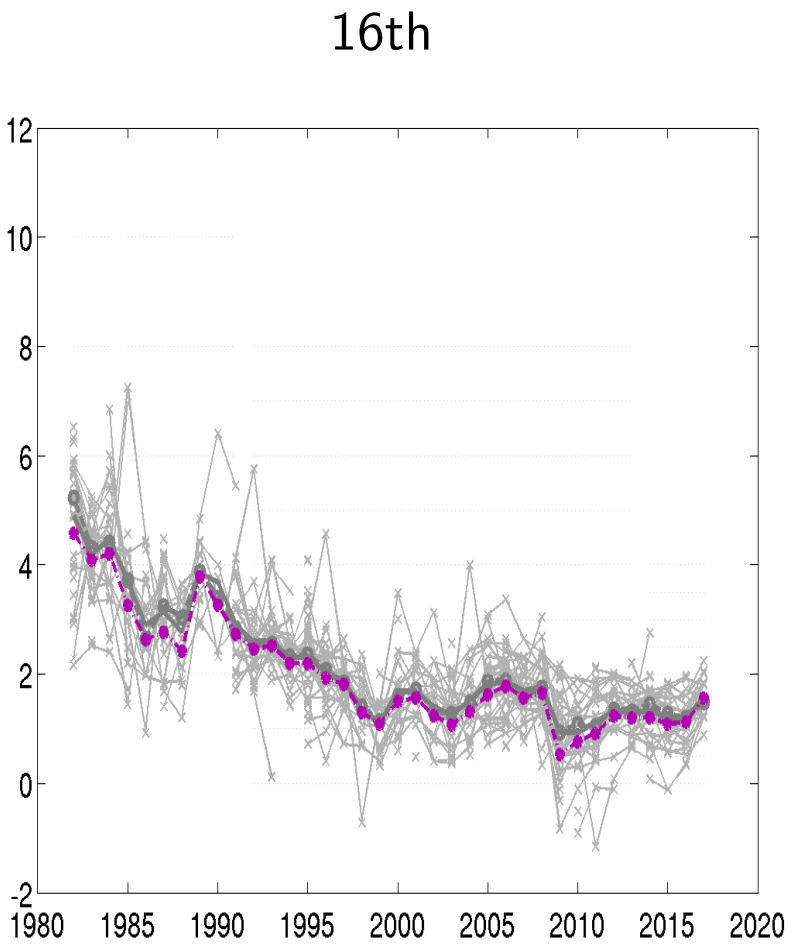
95th and 84th Quantiles





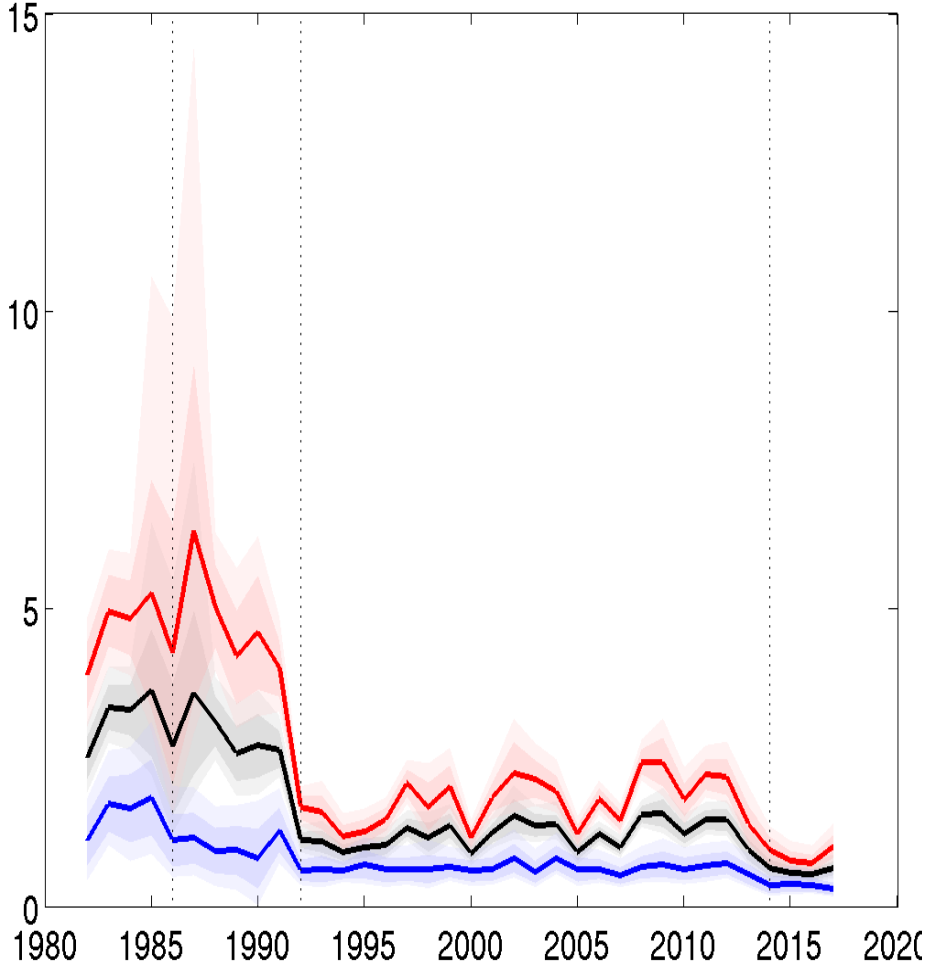
5th and 16th Quantiles



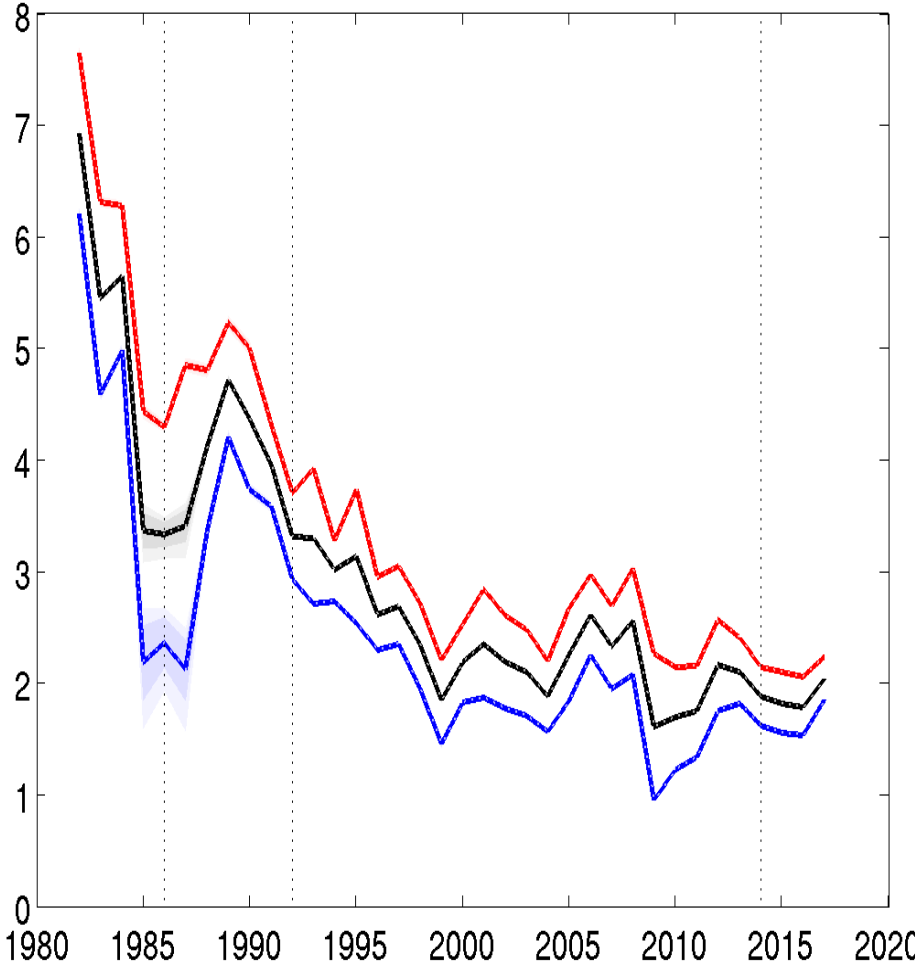


Heterogeneity in Variances and Means

Variance (Q1)

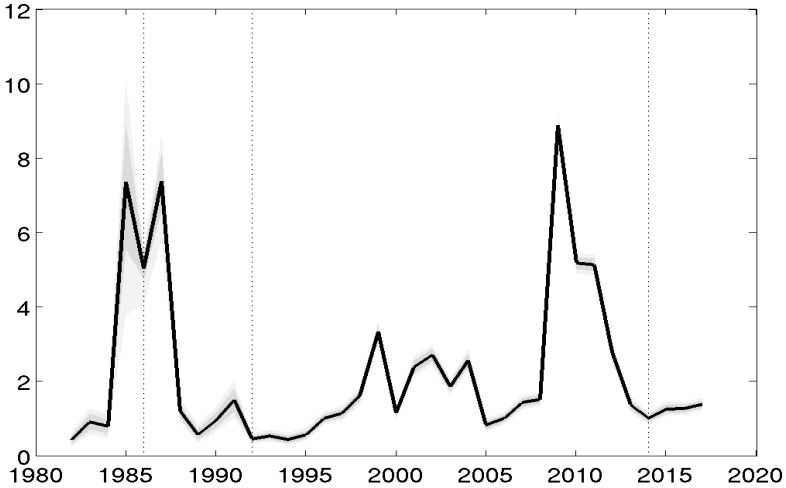


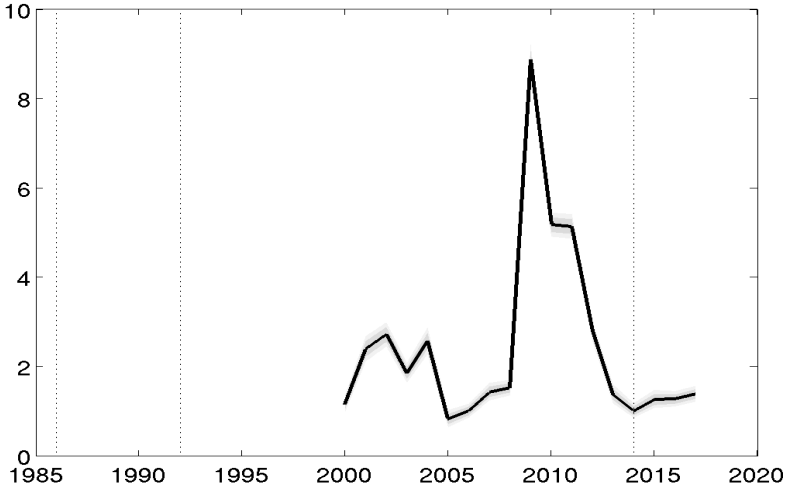
Mean (Q1)



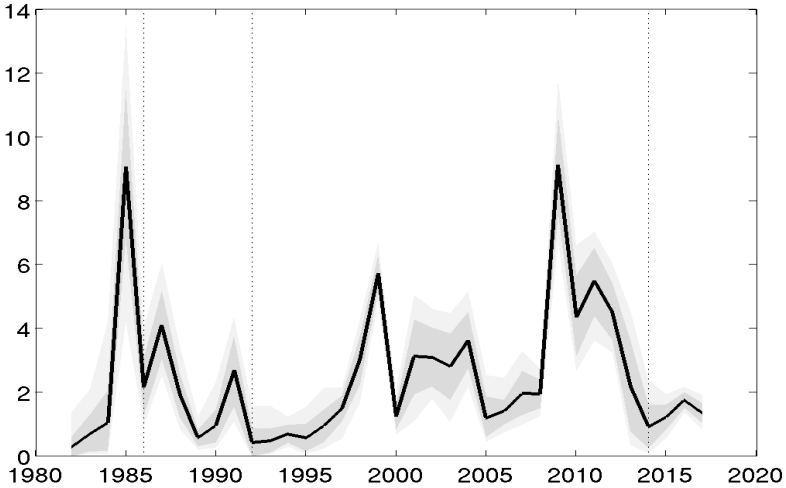
Extreme Events: Deflation Fear (H2Q1)

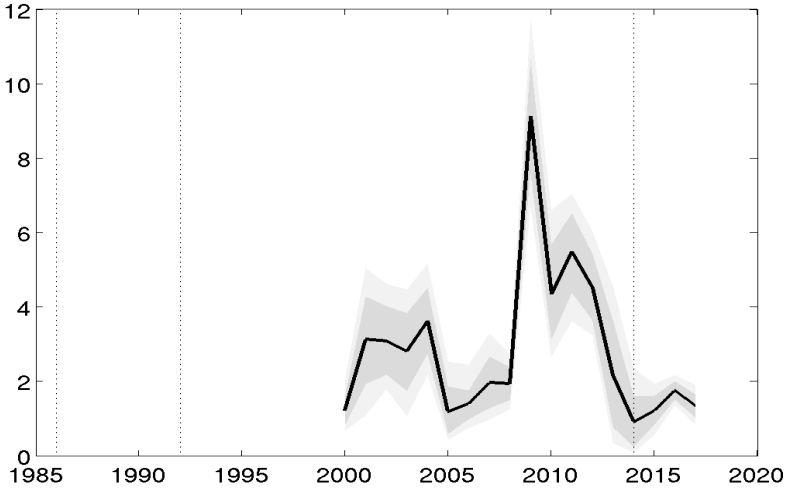
Average CDF





Top 25%





Asymptotic Properties: Posterior consistency

We apply **Schwartz's theorem**, which requires that for every $\varepsilon > 0$

$$\Pi(h : KL(h_0, h) > \varepsilon) > 0$$

where $KL(h_0, h)$ is the **Kullback-Leibler divergence** from h to h_0 defined as:

- $$KL(h_0, h) := \int h_0(\mathbf{z}) \log\left(\frac{h_0(\mathbf{z})}{h(\mathbf{z})}\right) d\mathbf{z}$$

when h_0 and h are continuous (i.e. no zeros is reported), and

- $$KL(h_0, h) := \sum_{\boldsymbol{\xi} \in \mathcal{X}} h_0(\boldsymbol{\xi}) \int_{\Delta^{J-|\boldsymbol{\xi}|}} h_0(\tilde{\mathbf{z}}|\boldsymbol{\xi}) \log\left(\frac{h_0(\tilde{\mathbf{z}}|\boldsymbol{\xi}) h_0(\boldsymbol{\xi})}{g(\tilde{\mathbf{z}}|\boldsymbol{\xi}) g(\boldsymbol{\xi})}\right) d\tilde{\mathbf{z}}$$

when h_0 and h are **discrete-continuous (mixed) distributions (i.e. some zeros are reported)**.

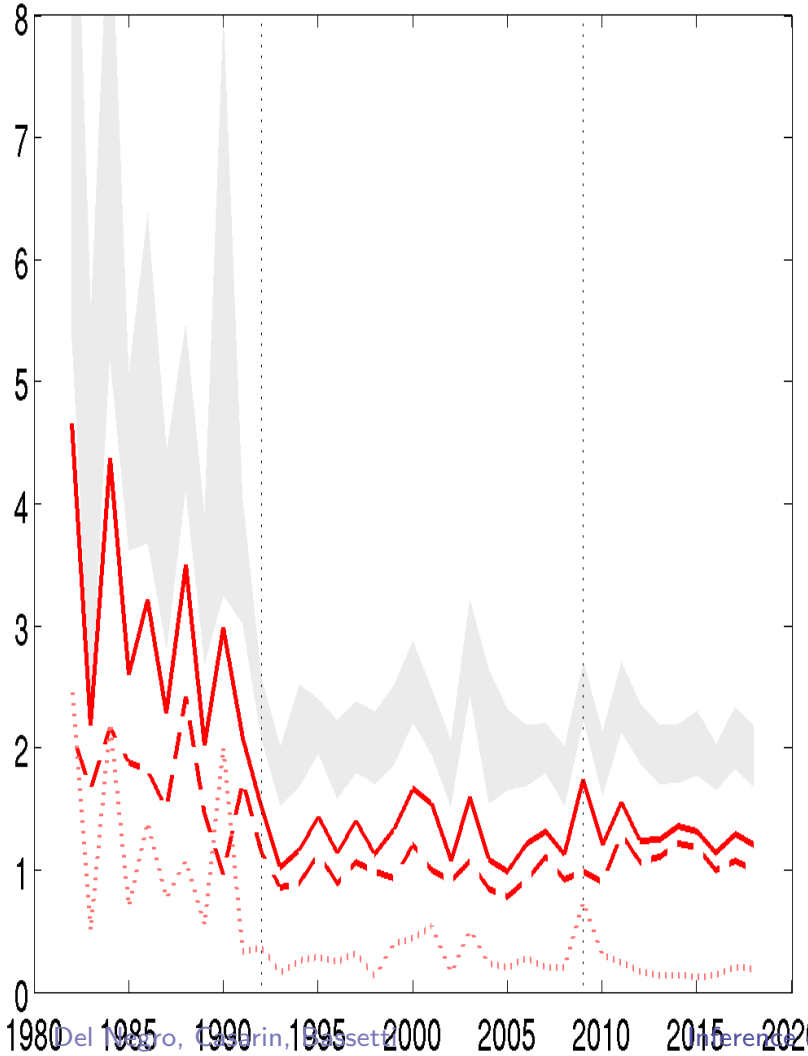
Notes:

- Consistency may depend on the properties of true density h_0.
- KL induces a topology stronger than the weak topology.

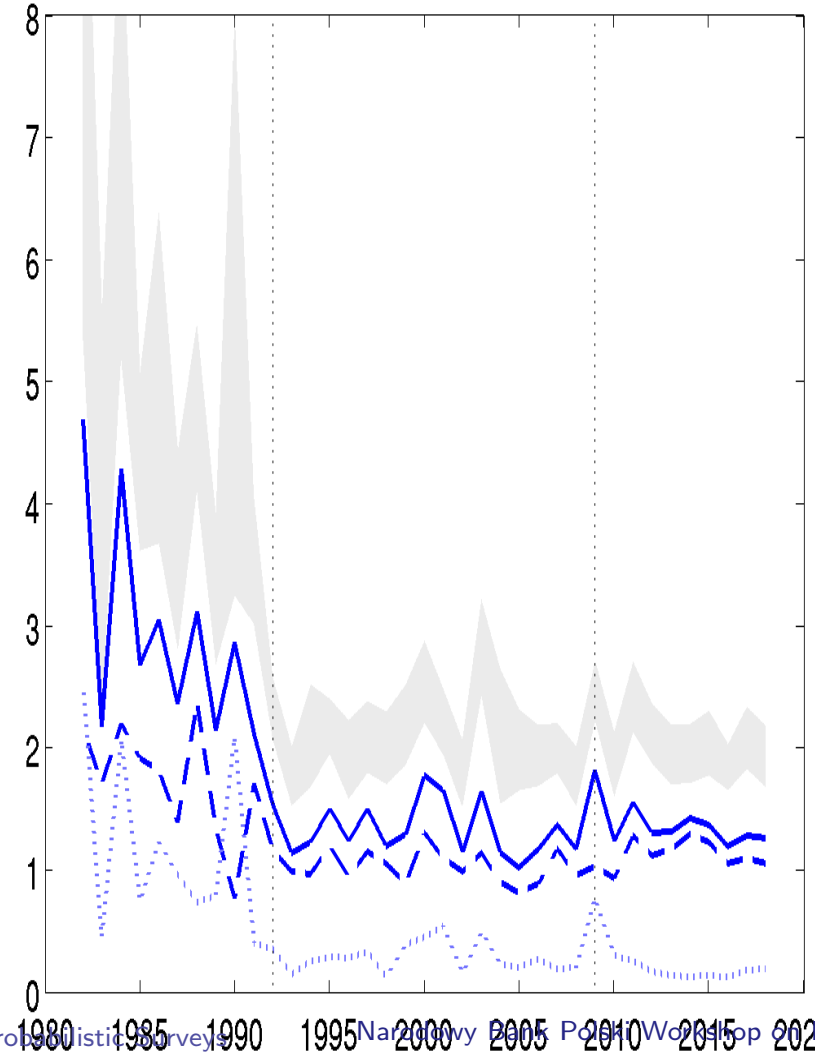
Variance, Uncertainty, and Disagreement: Alt. Methods (Output)

Output (H2Q2)

Beta

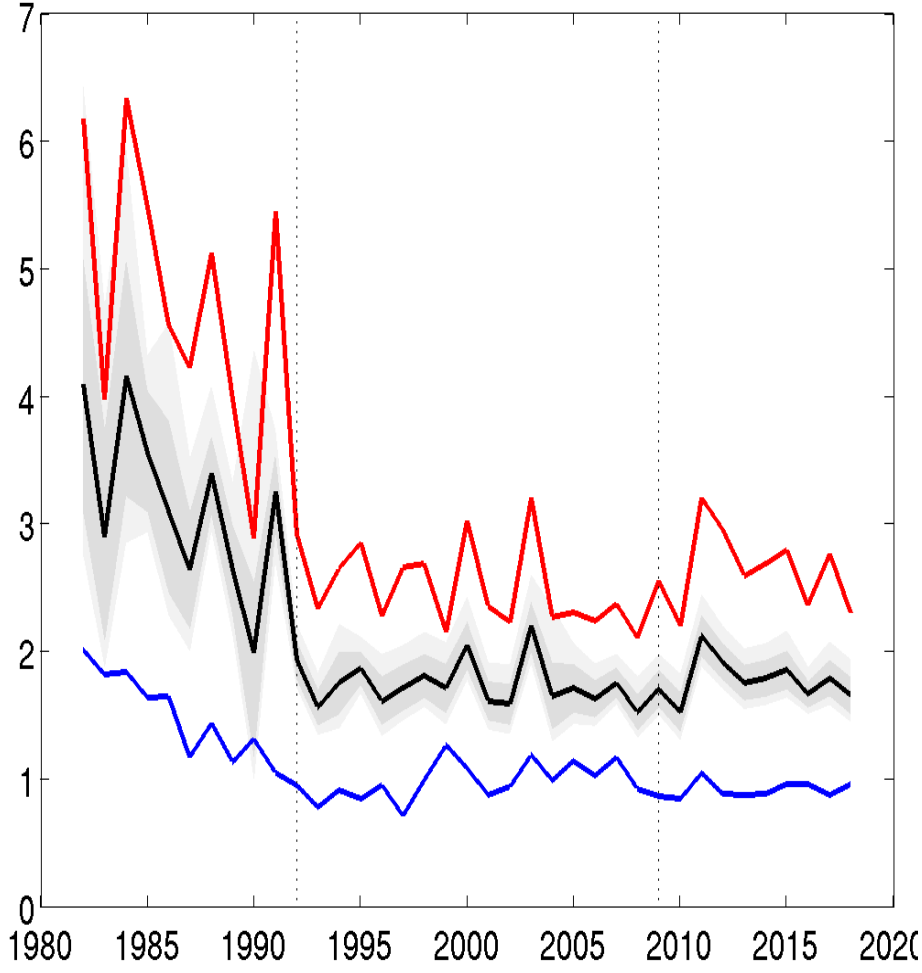


Normal

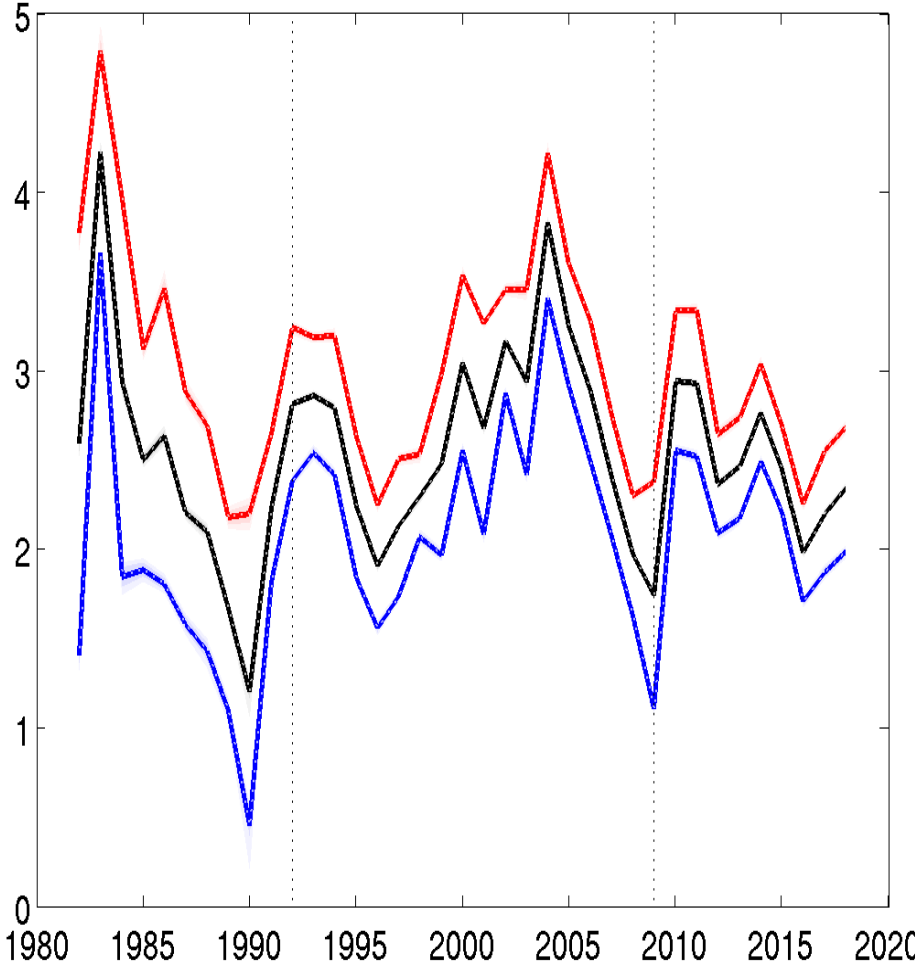


Heterogeneity in Means and Variances (Output)

Variance – High vs Low



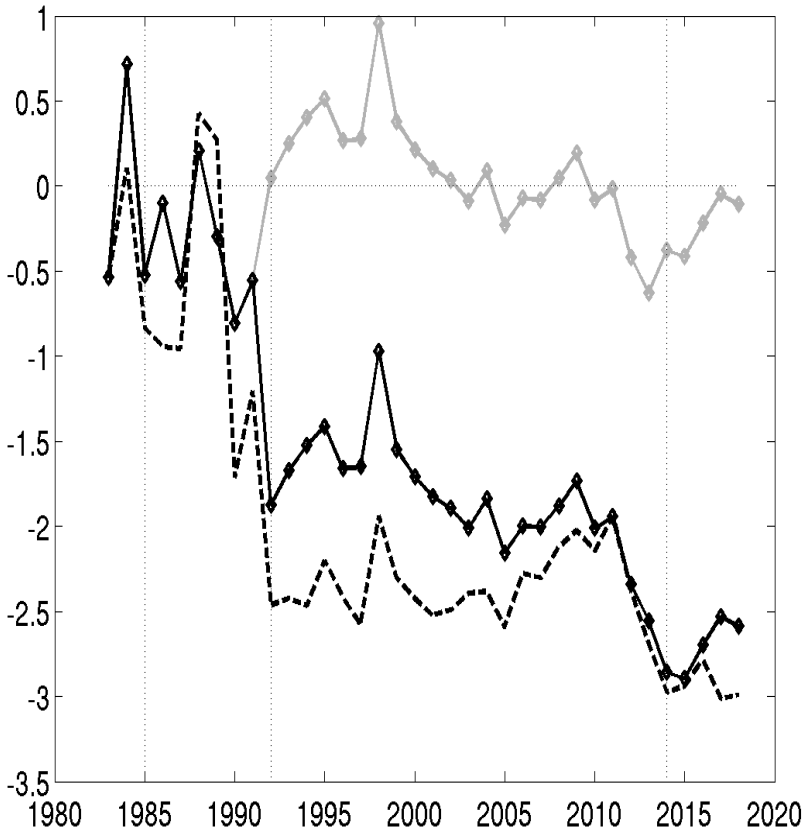
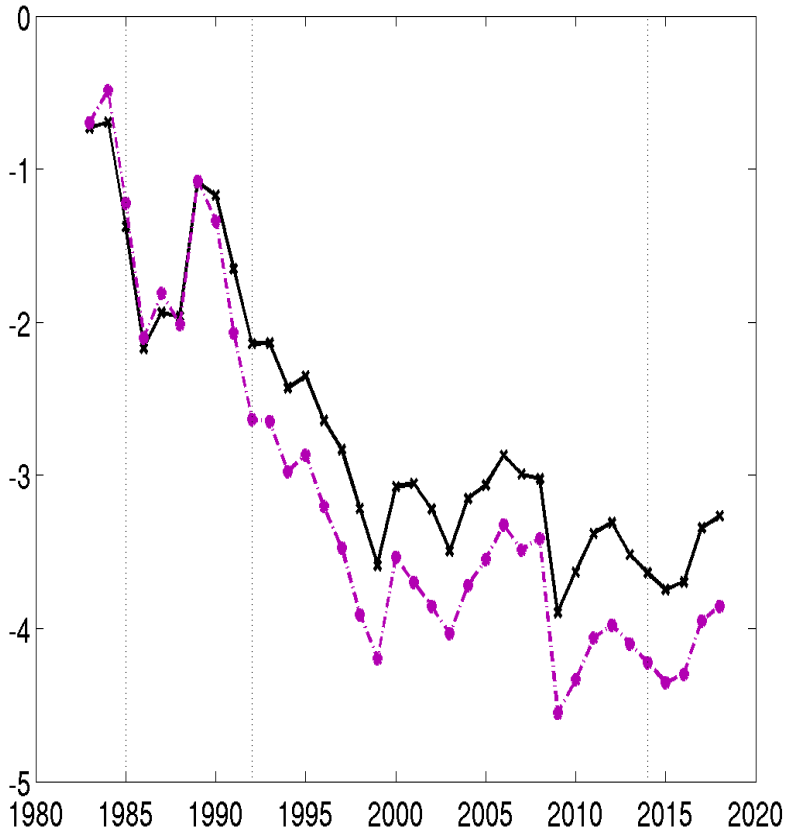
Mean – High vs Low



The Role of Survey Composition and Design

Cumulated Changes in

Variance



Mean