# From fixed-event to fixed-horizon density forecasts: professional forecasters' view on multi-horizon uncertainty

Gergely Ganics
*Banco de España*

Barbara Rossi
*ICREA–UPF,*
*Barcelona GSE, CREI*

Tatevik Sekhposyan
*Texas A&M University*

November 16, 2018

2nd "Forecasting at Central Banks" Conference
Bank of England

# Motivation

- Central banks: growing emphasis on communicating uncertainty around predictions (Galbraith and van Norden, 2012; Reifschneider and Tulip, 2017).
- Surveys of professional forecasters:
  1. precise and timely point forecasts for key macroeconomic variables (Ang et al., 2007; Del Negro and Schorfheide, 2013),
  2. but their *probabilistic* forecasts are less frequently used (Zarnowitz and Lambros, 1987; D'Amico and Orphanides, 2008; Clements, 2014b, 2018; Rossi and Sekhposyan, 2017).
  3. This is partly due to their fixed-event nature:
     - in each quarter, panelists forecast GDP growth and inflation in the current and the next **calendar** year,
     - forecast horizon contracts over time,
     - limiting usefulness for policy-makers and market participants.

**Contribution**

We combine fixed-event density forecasts into fixed-horizon density forecasts.

# Contribution & overview of results

1. Optimally weighted combination of fixed-event (current and next calendar year) density forecasts based on the US Survey of Professional Forecasters (between 1981Q3 and 2017Q2), resulting in four-quarter-ahead predictions between 1998Q3 and 2018Q1.

# Contribution & overview of results

1. Optimally weighted combination of fixed-event (current and next calendar year) density forecasts based on the US Survey of Professional Forecasters (between 1981Q3 and 2017Q2), resulting in four-quarter-ahead predictions between 1998Q3 and 2018Q1.
2. GDP growth and inflation: correctly calibrated density forecasts showing correct coverage.
   - Benchmarks:
     - mixture using ad-hoc weights,
     - Bayesian VAR,
     - model based on past forecast errors and SPF point forecasts.

# Contribution & overview of results

1. Optimally weighted combination of fixed-event (current and next calendar year) density forecasts based on the US Survey of Professional Forecasters (between 1981Q3 and 2017Q2), resulting in four-quarter-ahead predictions between 1998Q3 and 2018Q1.
2. GDP growth and inflation: correctly calibrated density forecasts showing correct coverage.
   - Benchmarks:
     - mixture using ad-hoc weights,
     - Bayesian VAR,
     - model based on past forecast errors and SPF point forecasts.
3. Investigating how to convert forecasters' histograms into continuous distributions: normal, Jones and Faddy's (2003) skew $t$ and Azzalini and Capitanio's (2003) skew $t$.
   - While skew $t$ distributions are advantageous particularly during the Great Recession, the choice of distribution seems to matter little for the final results.

# Roadmap

1. Related literature
2. Econometric framework
3. Empirical application: US SPF
   - Data
   - Results
4. Conclusion

# Related literature

The US SPF has been analyzed in various ways:

- Uncertainty: Giordani and Söderlind (2003), D'Amico and Orphanides (2008), Clements (2014a), Rossi et al. (2017), Clark et al. (2017).
- Evaluation: Rossi and Sekhposyan (2017), Clements (2018).

# Related literature

The US SPF has been analyzed in various ways:

- Uncertainty: Giordani and Söderlind (2003), D'Amico and Orphanides (2008), Clements (2014a), Rossi et al. (2017), Clark et al. (2017).
- Evaluation: Rossi and Sekhposyan (2017), Clements (2018).

Combining fixed-event *point* forecasts into fixed-horizon ones:

- Ad-hoc weights (Dovern et al., 2012).
- Minimizing expected squared error (Knüppel and Vladu, 2016).

# Related literature

The US SPF has been analyzed in various ways:

- Uncertainty: Giordani and Söderlind (2003), D'Amico and Orphanides (2008), Clements (2014a), Rossi et al. (2017), Clark et al. (2017).
- Evaluation: Rossi and Sekhposyan (2017), Clements (2018).

Combining fixed-event *point* forecasts into fixed-horizon ones:

- Ad-hoc weights (Dovern et al., 2012).
- Minimizing expected squared error (Knüppel and Vladu, 2016).

**How is our paper different?**

1. Q: What is the optimal way to combine fixed-event density forecasts into fixed-horizon ones?
2. A: Estimate weights based on the uniformity of the Probability Integral Transform (Ganics, 2017).

# Econometric framework I - What are fixed-event forecasts?

- In quarter $q$, survey panelists submit two fixed-event predictive distributions (histograms): $\widehat{F}^0_{t,q}(y)$ and $\widehat{F}^1_{t,q}(y)$, corresponding to current year's and next year's GDP growth (inflation).

| Quarter | Current Year $\widehat{F}^0_{t,q}(y)$ | Next Year $\widehat{F}^1_{t,q}(y)$ |
|---|---|---|
| $q = 1$ | $h = 4$ | $h = 8$ |
| $q = 2$ | $h = 3$ | $h = 7$ |
| $q = 3$ | $h = 2$ | $h = 6$ |
| $q = 4$ | $h = 1$ | $h = 5$ |

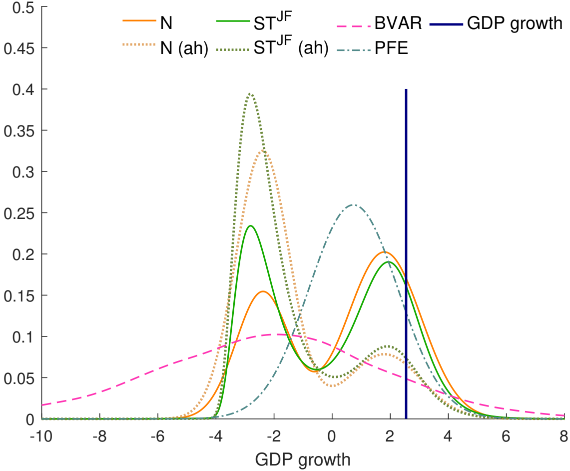
- We combine these distributions to obtain the "best" $h$-period-ahead forecast – "best" in a specific way.
- The combined CDF is in the class of linear opinion pools:

$$\widehat{F}^{q+h,c}_{t,q}(y) = w^h_{q,0}\widehat{F}^0_{t,q}(y) + w^h_{q,1}\widehat{F}^1_{t,q}(y) \qquad (1)$$

s.t. $0 \leq w^h_{q,0}, w^h_{q,1} \leq 1,\ w^h_{q,0} + w^h_{q,1} = 1,\ q \in \{1, 2, 3, 4\}$.

- Mixture structure provides flexibility.

Figure: Comparison of predictive densities for GDP growth as of 2009Q2



- A density forecast is probabilistically calibrated *iff* its Probability Integral Transform (PIT) is uniformly distributed (e.g. Corradi and Swanson (2006)).
- The estimator proposed by Ganics (2017) minimizes the discrepancy between the empirical CDF of the combined PIT and the uniform CDF in the Anderson–Darling sense.
- The PIT is the combined CDF evaluated at the realization:

$$\mathrm{PIT}_{t,q}^{q+h} \equiv \widehat{F}_{t,q}^{q+h,c}(y_{t,q}^{q+h}) \tag{2}$$

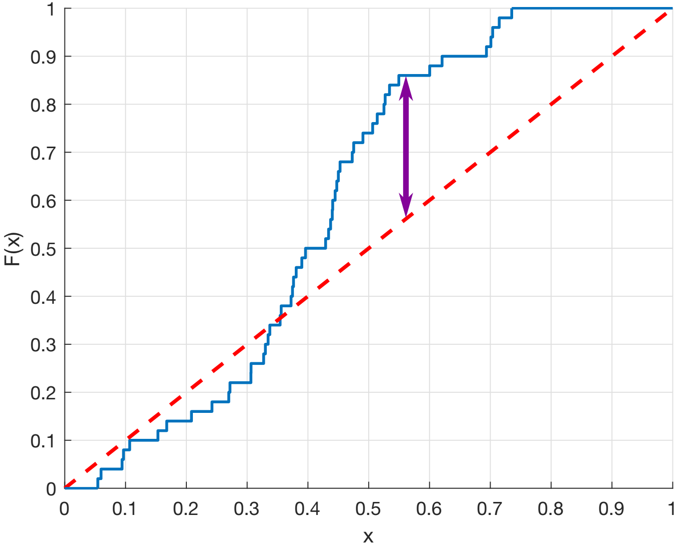
$$= w_{q,0}^{h}\widehat{F}_{t,q}^{0}(y_{t,q}^{q+h}) + w_{q,1}^{h}\widehat{F}_{t,q}^{1}(y_{t,q}^{q+h})\,. \tag{3}$$

- Vertical difference between the empirical CDF of the PITs and the uniform CDF at quantile $r \in [0,1]$:

$$\Psi_s(r, w^h) \equiv R^{-1} \sum_{t=s-R+1}^{s} \mathbb{1}\left[\mathrm{PIT}_{t,q}^{q+h} \leq r\right] - r\,, \tag{4}$$

where $s$ is the endpoint of a rolling window of $R$ observations, and $\mathbb{1}[\cdot]$ is the indicator function.

Figure: Visual example of $\Psi_s(r, w^h)$



# Econometric framework III

- To handle the small sample size and obtain truly out-of-sample forecasts, we parametrize the weights as exponential Almon lag polynomials (Andreou et al., 2010):

$$w^h_{q,0} \equiv B(\theta_1, \theta_2, q) \text{ for } q = 1, 2, 3, 4\,, \tag{5}$$

$$B(\theta_1, \theta_2, q) \equiv \exp(\theta_1 q + \theta_2 q^2)\,. \tag{6}$$

- We estimate weights through a modified version of the Anderson–Darling-type weight estimator of Ganics (2017):

$$\widehat{w}^h_{q,0} \equiv B(\widehat{\theta}_1, \widehat{\theta}_2, q)\,, \tag{7}$$

$$(\widehat{\theta}_1, \widehat{\theta}_2)' \equiv \underset{\theta_1, \theta_2 \in \Theta}{\operatorname{argmin}} \int_0^1 \frac{\Psi^2_s(r, w^h)}{r(1-r)}\, \mathrm{d}r\,, \tag{8}$$

where $\Theta$ is such that it ensures that the weights are positive, less than or equal to 1, and non-increasing in $q$.

# Data I

- We construct four-quarter-ahead density forecasts of quarterly year-on-year US **real GDP growth** and **inflation** measured by the GDP deflator, based on the US SPF between 1981Q3 and 2017Q2.
- All data (SPF and realizations) from Philadelphia Fed's Real-Time Data Research Center.
- Panelists provide their probabilistic forecasts of the **growth rate of the average level** of real GDP and GDP deflator from the previous calendar year to the current calendar year, and from the current calendar year to the next calendar year.
- These predictions take the form of probabilities assigned to pre-specified bins $\Rightarrow$ we transform them to continuous distributions.
- "Consensus": average across forecasters in each bin (*unlike* Del Negro et al., 2018).

# Data II

- In each quarter, we fit 3 distributions to current year's and next year's GDP growth and inflation forecast
  1. normal,
  2. Jones and Faddy's (2003) skew $t$,
  3. Azzalini and Capitanio's (2003) skew $t$,

  using NLLS:

$$\widehat{\theta}^d = \underset{\theta^d \in \Theta^d}{\text{argmin}} \sum_{i=1}^{S} \Big( F^d(s_i;\theta^d) - F(s_i) \Big)^2 , \qquad (9)$$

  where $F^d(s_i;\theta^d)$ is the CDF of distribution $d$, $F(s_i)$ is the cumulative histogram, and $s_i$ are the endpoints of the bins.
- This procedure provides sequences of predictive CDFs $\widehat{F}^0_{t,q,d}(y)$ and $\widehat{F}^1_{t,q,d}(y)$ $\Rightarrow$ inputs of the PIT-based weight estimator.

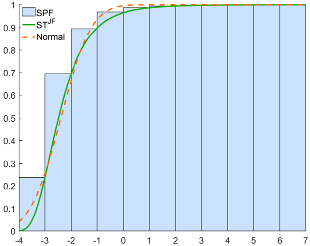
# Approximating the SPF distribution

Jones and Faddy's (2003) skew $t$ distribution

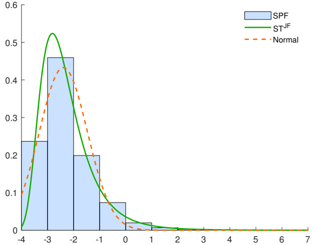
- Parametrization:
  1. skewness and tail behavior regulated by $a, b > 0$,
  2. location and a scale parameter $\mu$ and $\sigma > 0$.
- Includes Student's $t$ distribution as a special case, normal as a limiting case, it can display fat tails.
- Its CDF can be evaluated very quickly (regularized incomplete beta function).

# Figure: GDP growth: fitting normal and skew $t$ CDFs in 2009Q2

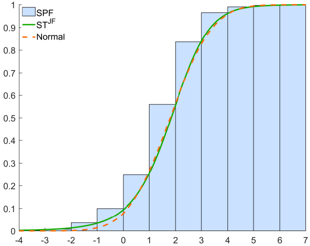
(a) CDF, current year




(b) PDF, current year




(c) CDF, next year




(d) PDF, next year

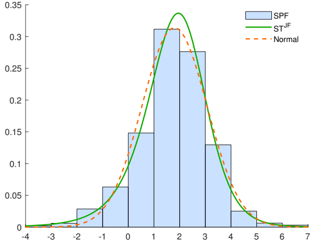
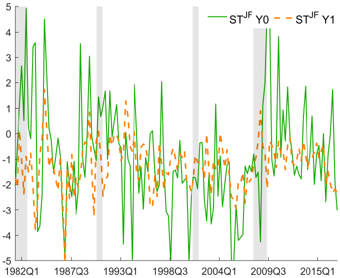
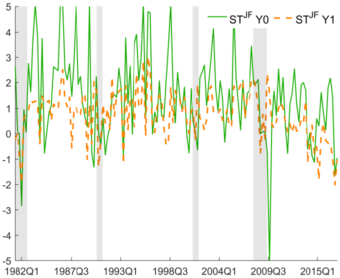


Figure: Skewness of fitted distributions between 1981Q3 and 2017Q2

(a) GDP growth




(b) Inflation




*Note*: Dates correspond to US SPF survey rounds. Shaded areas are NBER recession periods.

- Current year's forecasts more skewed than next year's forecasts.
- GDP growth forecasts mostly negatively skewed.
- Inflation forecasts usually positively skewed.

# Estimation I

Timing, example from the 1981Q3 survey:

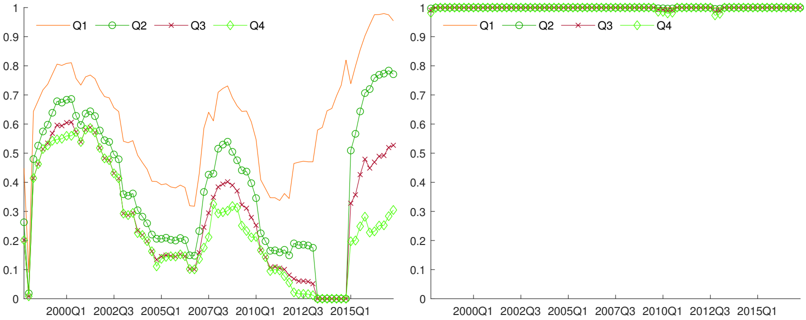
1. The quarter preceding the survey was 1981Q2, 4 quarters after that was 1982Q2.
2. First estimate of 1982Q2 GDP published in 1982Q3.
3. The percentage growth rate of the GDP estimates of 1982Q2 and 1981Q2 according to the 1982Q3 vintage is the first observation of real-time GDP growth.

# Estimation II

- Forecast origins: 1997Q4 to 2017Q2.
  1. At each forecast origin $s$, we take the most recent $R = 60$ survey forecasts $(\widehat{F}^0_{t,q}(y), \widehat{F}^1_{t,q}(y))$ and the corresponding GDP growth or inflation realizations $y^{q+h}_{t,q}$ based on the advance release, and evaluate the PITs.
  2. Form the Anderson–Darling-type objective function with the exponential Almon lag parametrization, and estimate the full weight vector $\Rightarrow \left\{\widehat{w}^h_{q,0}\right\}^4_{q=1}$.
  3. Depending on which quarter $q$ the forecast origin is, we combine the SPF forecasts corresponding to current and next years using estimated weight $\widehat{w}^h_{q,0}$ and $\widehat{w}^h_{q,1} = 1 - \widehat{w}^h_{q,0}$ to form the mixture distribution.
- Evaluation period: 1998Q3 to 2018Q1.

Figure: Weights on current year's density forecast

(a) GDP growth (b) Inflation

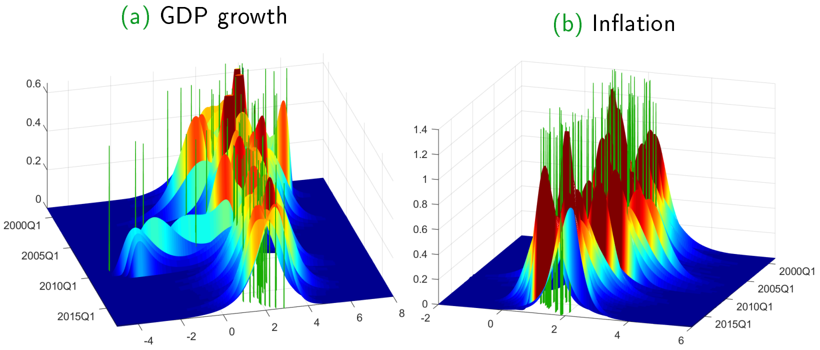



*Note*: The figures show the weight estimates at SPF survey rounds between 1997Q4 and 2017Q2.

- Weight estimates differ from ad-hoc weights (1,0.75,0.5,0.25).
- GDP growth: considerable time-variation.
- Inflation: all the weight on current year's forecast.

# Predictive densities

Figure: Four-quarter-ahead combined skew $t$ predictive densities




*Note*: The vertical green bars mark the realized values of the variable of interest based on the advance release. The forecast target dates on the horizontal axis range from 1998Q3 to 2018Q1.

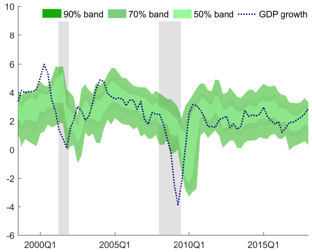
- GDP growth: striking skewness and bimodality around the Great Recession.
- Inflation: estimator "selects" current year's forecast $\Rightarrow$ tight predictive distributions.

# Benchmark models

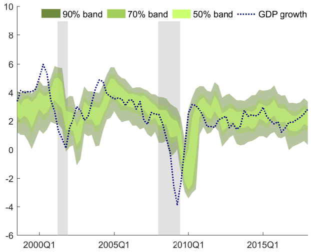
- This is not a forecasting horse race but we want to understand the properties of the proposed combined densities.
- We compared the performance of our proposed weighting scheme against:
  - an **ad-hoc** weighting method (Dovern et al. (2012),Rossi et al. (2017)), where the weights are determined by the overlap of forecast periods (denoted by ah),
  - a **Bayesian VAR with SV** using GDP, GDP deflator, TB3M, unemployment rate (Clark and Ravazzolo, 2015) estimated in rolling windows of 60 observations,
  - a **naive model** assuming normal predictive densities (Clements, 2018), serving as a simple version of Clark et al.'s (2017) stochastic volatility model (denoted by PFE):
    - variance estimated by MSE, calculated based on the past 60 four-quarter-ahead forecast errors,
    - the SPF point forecast is taken as the mean.

# Figure: GDP growth: Bands of predictive distributions, 1998Q3 – 2018Q1

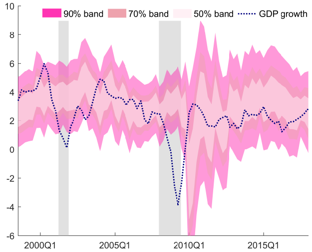
## (a) Optimal combination of skew $t$




## (b) Ad-hoc combination of skew $t$




## (c) BVAR




## (d) Past forecast errors

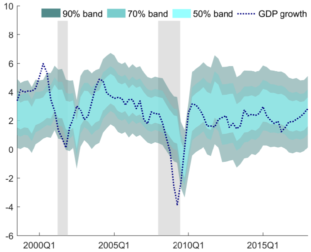
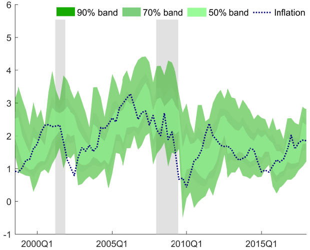
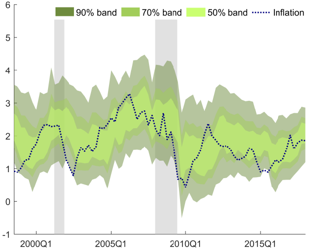


Figure: Inflation: Bands of predictive distributions, 1998Q3 – 2018Q1

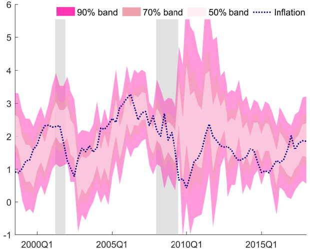
(a) Optimal combination of skew $t$



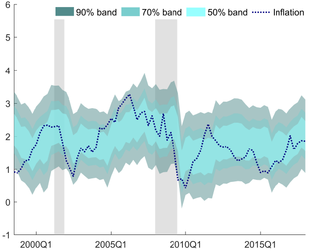

(b) Ad-hoc combination of skew $t$




(c) BVAR




(d) Past forecast errors



# Forecast evaluation: coverage rates

- Calculated the 25th and 75th percentiles (50% nominal rate), and the 15th and 85th percentiles (70% nominal rate) of the predictive distributions in each quarter.
- Calculated the ratio of cases when the realization of a particular variable fell inside ⇒ two-sided $t$ test.
- Correct coverage: optimally weighted SPF forecasts, BVAR.

Table: Absolute forecast evaluation: coverage

| | GDP growth | | Inflation | |
|---|---|---|---|---|
| | 50% | 70% | 50% | 70% |
| N | **49.4(0.92)** | **67.1(0.66)** | **55.7(0.36)** | **73.4(0.54)** |
| $ST^{JF}$ | **48.1(0.77)** | **63.3(0.31)** | **53.2(0.61)** | **73.4(0.52)** |
| N (ah) | **41.8(0.19)** | 57.0(0.06) | 63.3(0.04) | 81.0(0.03) |
| $ST^{JF}$(ah) | **41.8(0.19)** | 57.0(0.06) | 60.8(0.09) | 81.0(0.03) |
| BVAR | **55.7(0.39)** | **72.2(0.73)** | **55.7(0.35)** | **70.9(0.86)** |
| PFE | 65.8(0.02) | **77.2(0.22)** | 63.3(0.04) | **77.2(0.20)** |

*Note*: Empirical coverage rates, corresponding two-sided $p$-values of the null hypothesis of correct coverage in parentheses.

## Forecast evaluation: uniformity of PIT

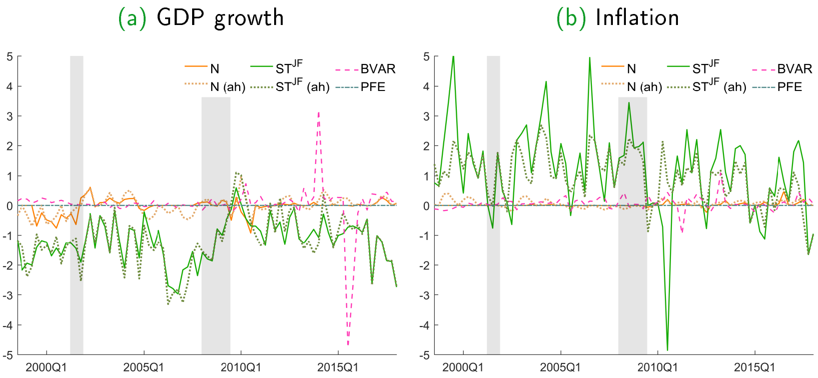
- Probabilistic calibration $\iff$ PIT uniformly distributed.
- Rossi and Sekhposyan's (2017) test of $H_0$ : PIT $\sim \mathcal{U}(0,1)$, using the Kolmogorov – Smirnov (KS) and Cramér–von Mises (CvM) test statistics.
- Our combination method is the only one which delivers correctly calibrated forecasts for both variables.

Table: Absolute forecast evaluation: PIT

| | GDP growth | | Inflation | |
|---|---|---|---|---|
| | KS | CvM | KS | CvM |
| N | **0.93(0.52)** | **0.19(0.60)** | **1.16(0.26)** | **0.40(0.23)** |
| $ST^{JF}$ | **0.94(0.51)** | **0.25(0.47)** | **0.91(0.47)** | **0.30(0.32)** |
| N (ah) | **0.96(0.48)** | **0.26(0.47)** | 1.68(0.06) | 0.90(0.05) |
| $ST^{JF}$(ah) | **1.03(0.40)** | **0.29(0.41)** | 1.71(0.05) | 0.82(0.06) |
| BVAR | **1.55(0.10)** | 0.80(0.08) | **0.90(0.51)** | **0.18(0.50)** |
| PFE | 1.72(0.06) | **0.62(0.12)** | **1.45(0.18)** | **0.53(0.17)** |

*Note*: Kolmogorov–Smirnov (KS) and Cramér–von Mises (CvM) test statistics, corresponding *p*-values in parentheses.

# Skewness

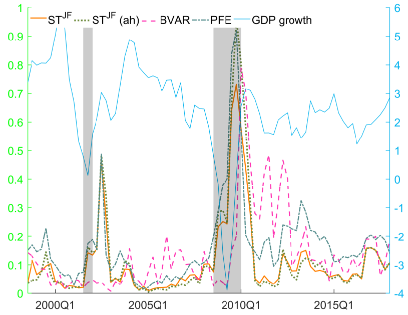



*Note*: For each model, the figures show the skewness of the predictive distributions defined as the standardized third central moment. The forecast target dates on the horizontal axis range from 1998Q3 to 2018Q1. Shaded areas are NBER recession periods.

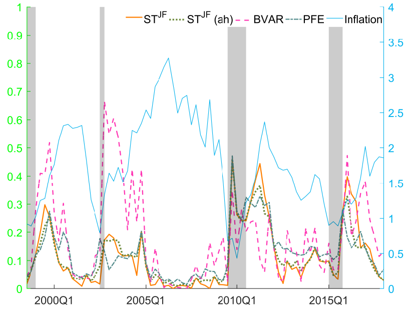
- Combinations of normals: practically no skewness.
- Combinations of skew $t$ distributions:
  1. GDP growth: negative skew.
  2. Inflation: mostly positive skew.

# Probabilities of adverse events

(a) Pr(GDP growth $\leq$ 1%)




(b) Pr(Inflation $\leq$ 1%)




*Note*: The figures show the probabilities of low GDP growth and low inflation (left axis) according to each model, along with the actual realization of the respective variable (solid blue line, right axis). The forecast target dates on the horizontal axis range from 1998Q3 to 2018Q1. Shaded areas are the periods when the predicted event (e.g. GDP growth $\leq$ 1%) did occur.

- GDP growth:
  - all methods pick up the Great Recession with a lag, and
  - combinations of surveys adapt faster after large shocks.
- Inflation: timing is better than for GDP growth.

# Conclusion

- We provide a flexible and simple data-driven tool for practitioners to convert fixed-event density forecasts to fixed-horizon ones.
- Optimal combinations of SPF density forecasts for four-quarter-ahead GDP growth and inflation are correctly calibrated.
- Our method often outperforms popular benchmarks.
- The ad-hoc weighting scheme performs poorly $\Rightarrow$ not recommended in practice.
- Further topics
  1. tail events, risk measures (Adrian, Boyarchenko and Giannone, 2018),
  2. comparison with Clark, McCracken and Mertens (2018).

Thank you for your attention!

# Additional results

Azzalini and Capitanio's (2003) skew $t$ distribution

Jones and Faddy's (2003) skew $t$ distribution

CRPS results

# Azzalini and Capitanio's (2003) skew $t$ distribution

- Location parameter $\mu$, scale parameter $\sigma > 0$, skewness parameter $\alpha$ and degrees of freedom parameter $\nu > 0$.

Its PDF at $x \in \mathbb{R}$ is given by

$$f(x; \mu, \sigma, \alpha, \nu) = \frac{2}{\sigma} t_\nu \left( \frac{x-\mu}{\sigma} \right) T_{\nu+1} \left( \alpha \frac{x-\mu}{\sigma} \sqrt{\frac{\nu+1}{\nu + \left(\frac{x-\mu}{\sigma}\right)^2}} \right),$$

where

- $t_\nu(\cdot)$ is the PDF of Student's $t$ distribution with degrees of freedom parameter $\nu$, and
- $T_{\nu+1}(\cdot)$ is the CDF of Student's $t$ distribution with degrees of freedom parameter $\nu+1$.

To evaluate the CDF of Azzalini and Capitanio's (2003) skew $t$ distribution, the PDF must be integrated numerically.

Back

# Jones and Faddy's (2003) skew $t$ distribution

The distribution's PDF at $x \in \mathbb{R}$ is given by

$$f(x; \mu, \sigma, a, b) = \frac{1}{\sigma} C_{a,b}^{-1} \left(1+\tau\right)^{a+1/2} \left(1-\tau\right)^{b+1/2} \quad (10)$$

$$C_{a,b} = 2^{a+b-1} B(a,b)(a+b)^{\frac{1}{2}} \quad (11)$$

$$\tau = \frac{x-\mu}{\sigma}\left(a+b+\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{1}{2}}, \quad (12)$$

The distribution's CDF at $x \in \mathbb{R}$ is given by

$$F(x; \mu, \sigma, a, b) = I_z(a,b) \quad (13)$$

$$z = \frac{1}{2}\left(1 + \frac{\left(\frac{x-\mu}{\sigma}\right)}{\sqrt{a+b+\left(\frac{x-\mu}{\sigma}\right)^2}}\right), \quad (14)$$

where $I_v(\cdot,\cdot)$ is the regularized incomplete beta function (a.k.a. the incomplete beta function ratio).

Back

# Empirical application: CRPS

- Continuous Ranked Probability Score: strictly proper scoring rule (Gneiting and Raftery, 2007), frequently used in the literature to evaluate forecasts.
- For the *h*-quarter-ahead density forecast made in year *t* and quarter *q* using model *m*, it is defined as

$$\mathrm{CRPS}_{t,q+h}^{(m)} \equiv \int_{-\infty}^{\infty} \left( \widehat{F}_{t,q}^{q+h\,(m)}(y) - \mathbb{1}\left[ y_{t,q}^{q+h} \leq y \right] \right)^2 \mathrm{d}y \,, \quad (15)$$

where $\widehat{F}_{t,q}^{q+h\,(m)}(y)$ is the corresponding predictive CDF. The average full-sample CRPS is given by

$$\mathrm{CRPS}^{(m)} \equiv T^{-1} \sum_{q=1}^{4} \sum_{t=1}^{T_q} \mathrm{CRPS}_{t,q+h}^{(m)} \,. \quad (16)$$

- Lower values of the CRPS correspond to better models.

# Empirical application: relative predictive ability

- GDP growth: clear superiority of our combination method.
- Inflation: mixed results, but never significantly outperformed.

Table: Relative forecast evaluation: CRPS

| | GDP growth | Inflation |
|---|---|---|
| N | 0.75 | 0.34 |
| $ST^{JF}$ | **0.75** | 0.34 |
| N (ah) | 0.79 | 0.33 |
| $ST^{JF}$(ah) | 0.79 | 0.33 |
| BVAR | 1.05 | 0.42 |
| PFE | 0.76 | **0.32** |
| N vs N (ah) | $-0.99(0.16)$ | $0.98(0.84)$ |
| $ST^{JF}$ vs $ST^{JF}$(ah) | $-1.35(0.09)^{*}$ | $1.19(0.88)$ |
| N vs BVAR | $-3.42(0.00)^{***}$ | $-2.81(0.00)^{***}$ |
| $ST^{JF}$ vs BVAR | $-3.60(0.00)^{***}$ | $-2.74(0.00)^{***}$ |
| N vs PFE | $-0.16(0.44)$ | $0.74(0.77)$ |
| $ST^{JF}$ vs PFE | $-0.32(0.37)$ | $0.94(0.83)$ |

*Note*: Top panel: CRPS. Bottom panel: Diebold and Mariano (1995) test statistics, and $p$-values in parentheses, with rejection region in the left tail comparing predictive accuracy measured by CRPS. Negative value indicates the first method outperforms the second one, asterisks denote rejection.