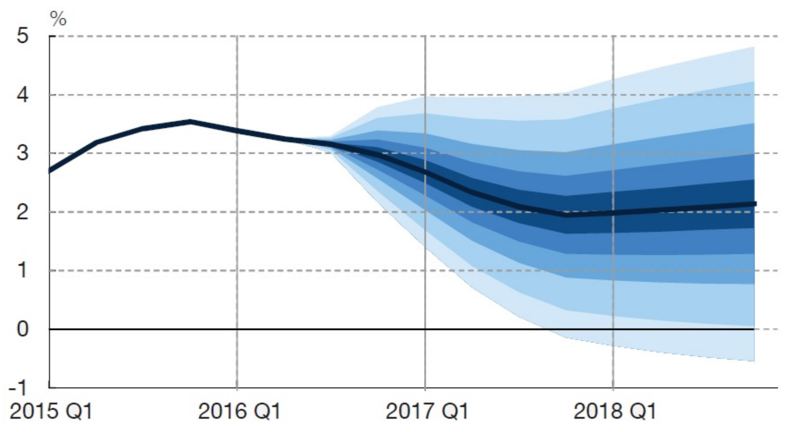


Optimal Density Forecast Combinations

GERGELY GANICS
gergely.ganics@bde.es

BANCO DE ESPAÑA
Eurosistema

MOTIVATION

- Density forecasts summarize *uncertainty* surrounding point forecasts, hence they facilitate communication between researchers and decision makers and to the general public (fan charts).
- Combining models' density forecasts has a similar motivation as in the point forecasting framework: mitigating model misspecification, parameter estimation uncertainty (Timmermann, 2006).
- Rich literature on combining point forecasts (Bates and Granger, 1969; Stock and Watson, 2004; Cheng and Hansen, 2015), much less on density forecasts (Hall and Mitchell, 2007; Geweke and Amisano, 2011; Rossi and Sekhposyan, 2014).
- Most of the literature focuses on *evaluating* density forecasts (Diebold et al., 1998; Corradi and Swanson, 2006; Rossi and Sekhposyan, 2013, 2016) and the combination schemes are often ad-hoc.

CONTRIBUTION

1. **Theory**: I propose a consistent estimator of combination weights that minimize the discrepancy between the combined density forecast and the probabilistically calibrated forecast.
2. **Monte Carlo**: the estimator performs well in finite samples.
3. **Empirics**: the combination scheme delivers probabilistically calibrated density forecasts when predicting US industrial production.

NOTATION

- Convex combination of $\mathcal{M}$ h-period-ahead predictive densities, conditional on information between $t-R+1$ and t:

$$\phi^C_{t+h}(y_{t+h}|\mathfrak{I}^t_R) \equiv \sum_{m=1}^{\mathcal{M}} w_m \phi^m_{t+h}(y_{t+h}|\mathfrak{I}^t_R)$$

- Probability Integral Transform (PIT) with realization Y_{t+h}:

$$\text{PIT}_{t+h} \equiv \int_{-\infty}^{Y_{t+h}} \phi^C_{t+h}(y|\mathfrak{I}^t_R)\,\text{d}y = \Phi^C_{t+h}(Y_{t+h}|\mathfrak{I}^t_R)$$

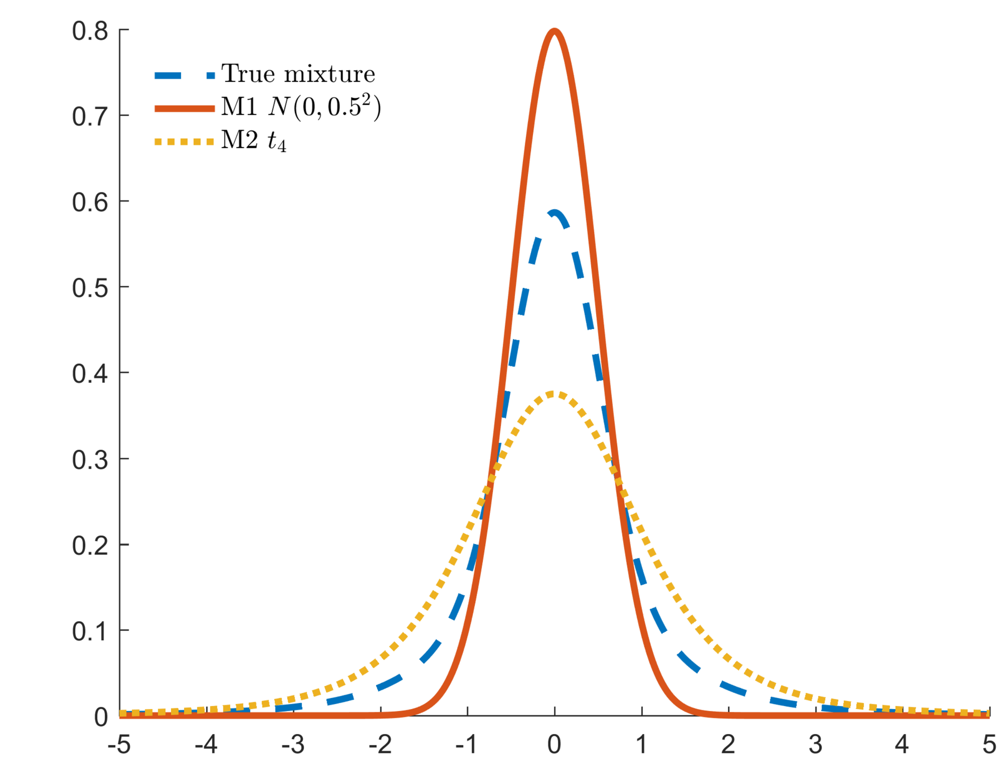
- True conditional distribution of y_{t+h}: $\phi^*_{t+h}(y_{t+h}|\mathfrak{I}^t_R)$.
- Probabilistic calibration (no reference to the true DGP!):

$$\phi^C_{t+h}(y_{t+h}|\mathfrak{I}^t_R) = \phi^*_{t+h}(y_{t+h}|\mathfrak{I}^t_R)$$

> $\text{PIT}_{t+h} \sim \mathcal{U}(0,1)$ *iff* probabilistic calibration holds.

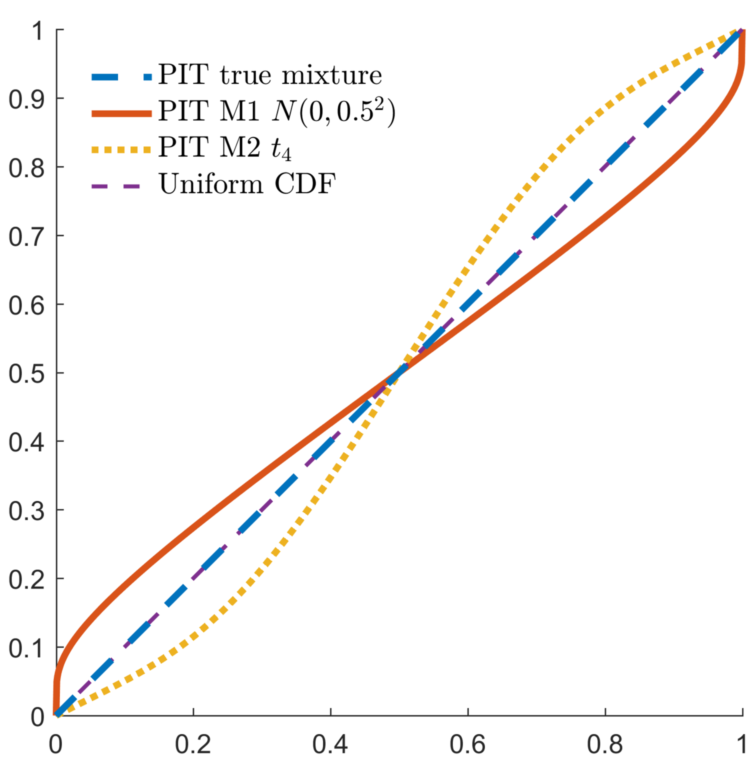
DETECTING (THE LACK OF) PROBABILISTIC CALIBRATION

Predictive densities



- True predictive density: $\phi^*_{t+1}(y_{t+1}|\mathfrak{I}^t_R) = 0.5\mathcal{N}(0,0.5^2) + 0.5t_4$
- Incorrect M1: $\phi^1_{t+1}(y_{t+1}|\mathfrak{I}^t_R) = \mathcal{N}(0,0.5^2)$
- Incorrect M2: $\phi^2_{t+1}(y_{t+1}|\mathfrak{I}^t_R) = t_4$

CDFs of Probability Integral Transforms



- Probabilistically calibrated forecast: CDF of PIT is the 45 degree line.
- M1, M2: markedly different tails show up in the CDFs of their PITs.

PROPOSED ESTIMATOR OF LINEAR COMBINATION WEIGHTS

- Idea: minimize the distance between the empirical CDF of the PIT and the CDF of the uniform distribution.
- Distance between uniform CDF and combined CDF at $r \in [0,1]$:

$$\Psi_G(r,w) \equiv G^{-1} \sum_{l=t-G+1-h}^{t-h} \mathbf{1}\,[\text{PIT}_{l+h} \leq r] - r$$

- Three well-known statistics, which differ in how they weigh vertical differences between the CDF's over the unit interval:

$$K_G(w) \equiv \sup_{r\in[0,1]} |\Psi_G(r,w)| \qquad \text{(Kolmogorov–Smirnov)}$$

$$C_G(w) \equiv \int_0^1 \Psi^2_G(r,w)\,\text{d}r \qquad \text{(Cramer–von Mises)}$$

$$A_G(w) \equiv \int_0^1 \Psi^2_G(r,w)\,[r(1-r)]^{-1}\,\text{d}r \qquad \text{(Anderson–Darling)}$$

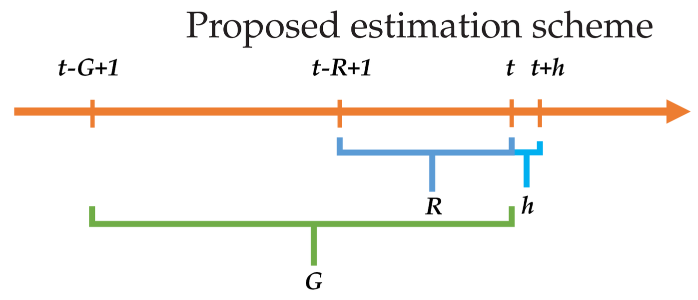
- Estimator:

$$\widehat{w} \in \Delta^{\mathcal{M}-1} \text{ s.t. } T_G(\widehat{w}) \leq \inf_{w\in\Delta^{\mathcal{M}-1}} T_G(w) + o_P(1),$$

where $T_G(w)$ is one of $K_G(w), C_G(w)$ or $A_G(w)$.

- Misspecification allowed: true conditional density does not need to belong to the span of the individual densities.

Proposed estimation scheme



> Estimator is **consistent** under mild mixing and continuity conditions.

MONTE CARLO EVIDENCE

- True DGP is a mixture of models M1 and M2 with mixture weights $(w_1, w_2)' = (0.4, 0.6)'$, while $M3$ is irrelevant, $w_3 = 0$:

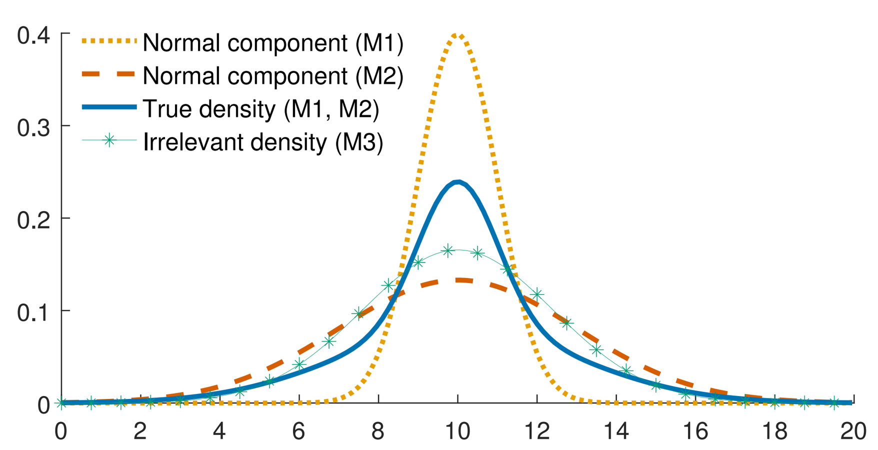
$$M1: y_{t+1} = 1 + 0.9y_t + \nu_{t+1}, \nu_{t+1} \overset{\text{iid}}{\sim} \mathcal{N}(0,1^2)$$

$$M2: y_{t+1} = 1 + 0.9y_t + \varepsilon_{t+1}, \varepsilon_{t+1} \overset{\text{iid}}{\sim} \mathcal{N}(0,3^2)$$

$$M3: y_{t+1} = 1 + 0.9y_t + \eta_{t+1}, \eta_{t+1} \overset{\text{iid}}{\sim} \mathcal{N}(0,5.8)$$

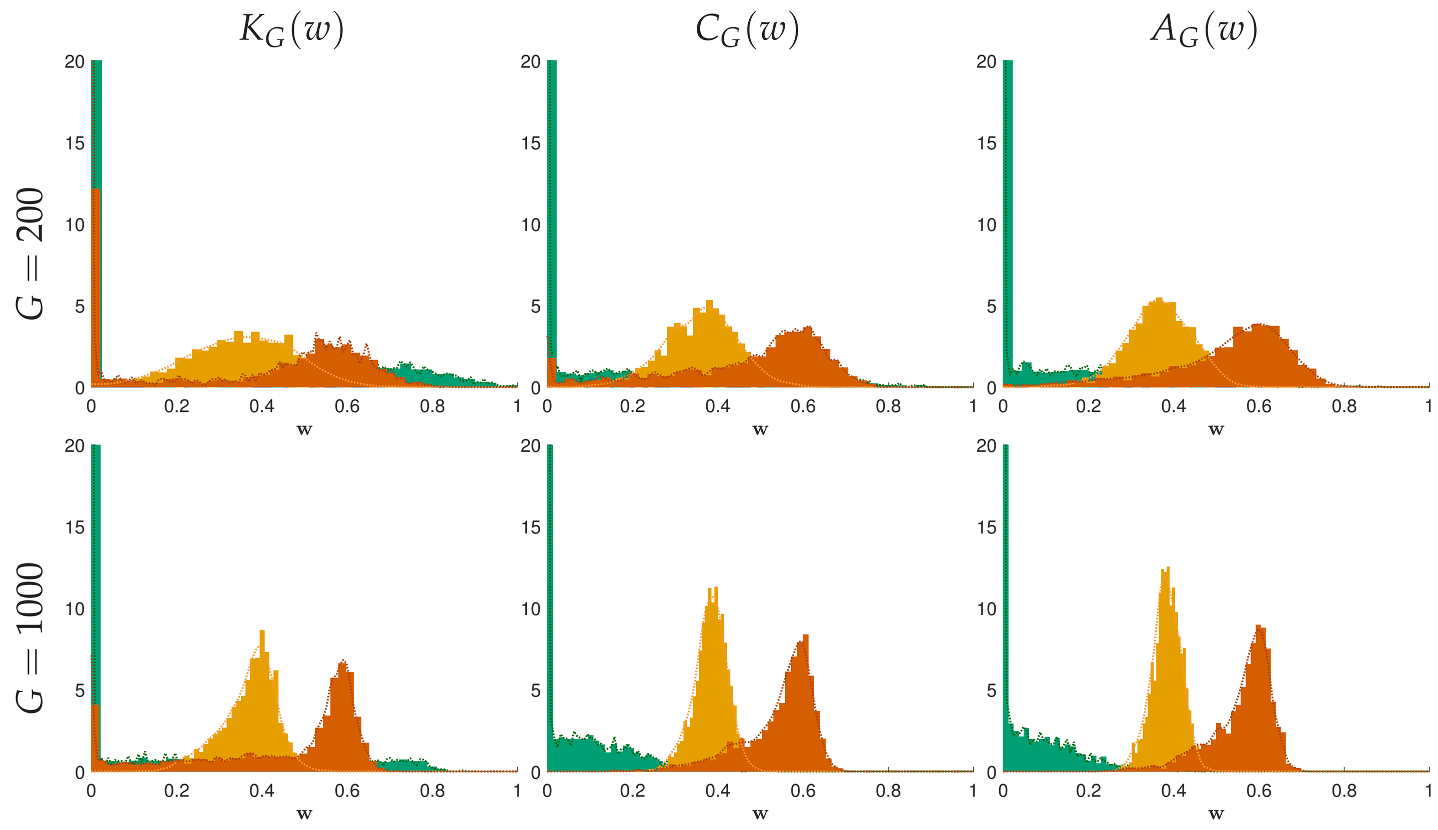
- First two moments of the true mixture density and the irrelevant density (M3) are the same but their shapes differ!
- Specifically, very different tails, implying vastly different predictions on the range of "extreme" future events.

Predictive densities of M1, M2, M3 and true mixture



Monte Carlo results

- Monte Carlo simulations using all three objective functions $T_G(w) = \{K_G(w), C_G(w), A_G(w)\}$ and sample sizes $G = \{200, 1000\}$.



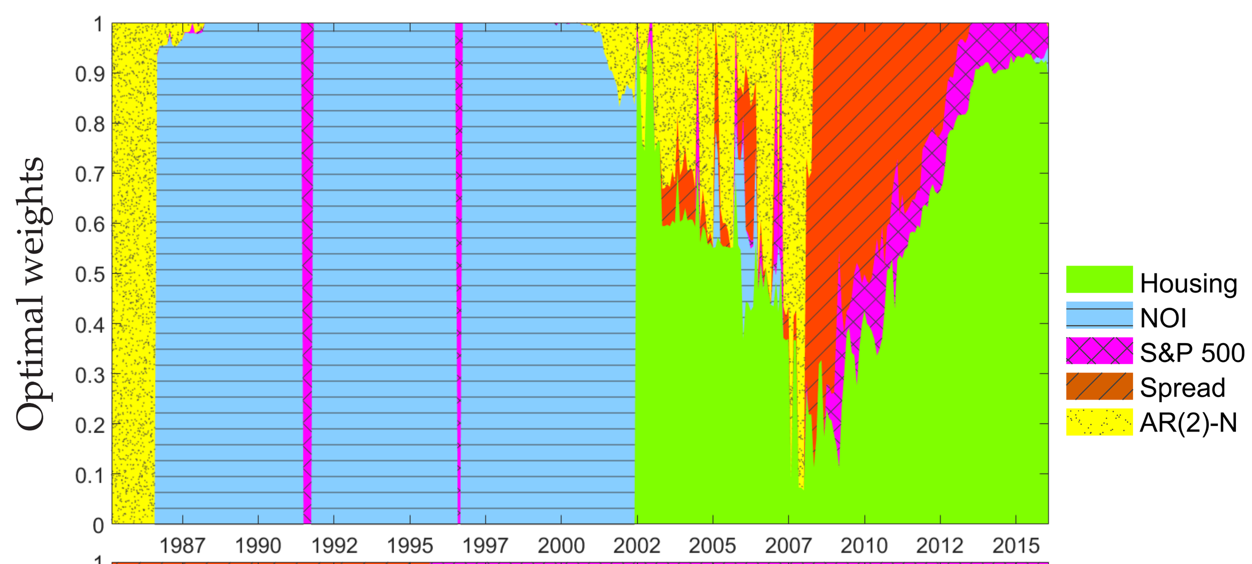
> KS-based estimator performs poorly, while the estimators based on the CvM and the **Anderson–Darling** statistics perform very well, the latter dominating the other two in the MSE sense.

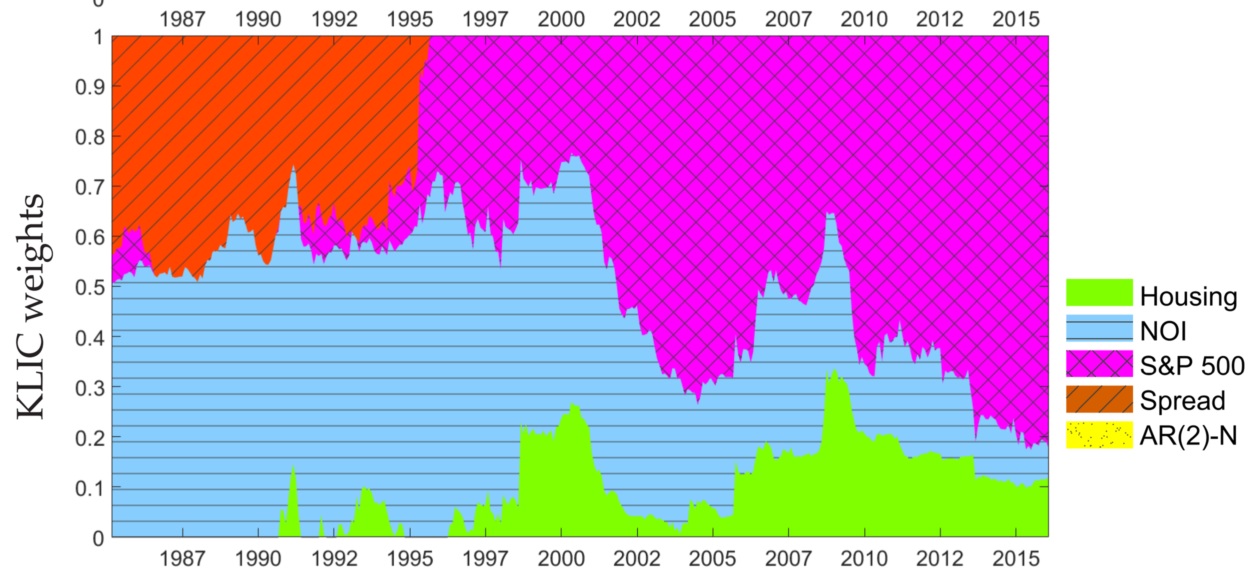
EMPIRICAL APPLICATION: FORECASTING US INDUSTRIAL PRODUCTION ONE MONTH AHEAD

- Autoregressive Distributed Lags (ARDL) models:

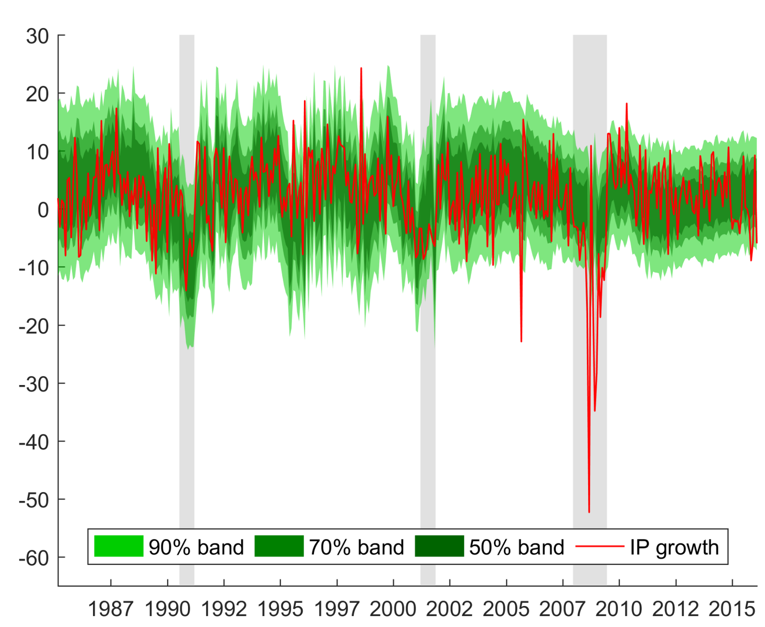
$$\text{IP}_{t+1} = c + \sum_{j=0}^{1} \beta_j \text{IP}_{t-j} + \sum_{j=0}^{1} \gamma_j X_{t-j} + \sigma\varepsilon_{t+1}, \ \varepsilon_{t+1} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$$

- X_t ={Capacity Util. in Mfg., ISM: New Orders Index, S&P's 500, Moody's Baa spread,∅}, one at a time, from FRED-MD.
- Each model estimated by ML in rolling windows of $R = 120$ months, forecast target dates 1985:03 – 2016:02 ($P = 372$ months).
- Weights estimated using $G = 180$ data points, Anderson–Darling-type objective function A_G.
- Benchmarks:
  - equal weights (Kascha and Ravazzolo, 2010),
  - maximizing the in-sample log scores, minimizing the KLIC between the true forecasting density and the combination density (Hall and Mitchell, 2007),
  - AR(2) with normal errors (Del Negro and Schorfheide, 2013).





Combined density forecasts using optimal weights



Testing the null hypothesis of uniformity:
Rossi and Sekhposyan (2016) test statistics (p-values)

	KS	CvM
Optimal weights	0.90 (0.38)	0.24 (0.22)
Equal weights	1.39 (0.05)	0.50 (0.04)
KLIC weights	1.28 (0.08)	0.40 (0.09)
AR(2)-N	1.31 (0.08)	0.62 (0.02)

> New weight estimation scheme delivers **probabilistically calibrated** density forecasts, **beats equal weighting**. Importance of **financial variables** during/after the Great Recession, extending Ng and Wright (2013).