Optimal Density Forecast Combinations

Gergely Gánics

Banco de España

gergely.ganics@bde.es

13th Annual Conference on Real-Time Data Analysis, Methods and Applications
Banco de España

October 19, 2017

Disclaimer: The views expressed herein are those of the author and should not be attributed to the Banco de España or the Eurosystem.

What is a density forecast?

- Density forecasts provide a probabilistic description of uncertainty surrounding forecasts.
- Q: Where do we use them?

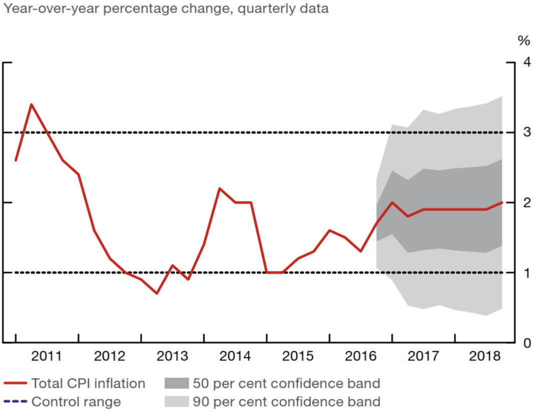
What is a density forecast?

- Density forecasts provide a probabilistic description of uncertainty surrounding forecasts.
- Q: Where do we use them?
- A: Weather forecasts, traffic forecasts, economics, ...

What is a density forecast?

- Density forecasts provide a probabilistic description of uncertainty surrounding forecasts.
- Q: Where do we use them?
- A: Weather forecasts, traffic forecasts, economics, . . .

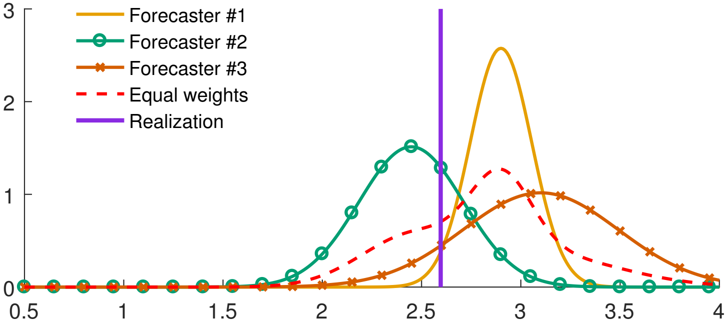
Figure: October 2016 Monetary Policy Report, Bank of Canada



The problem of density forecast combinations

- Survey of Professional Forecasters: large-scale, quarterly macroeconomic survey maintained by the Philadelphia Fed.

Figure: Forecasters' predictions for US ΔGDP in 2015 as of 2015:Q1



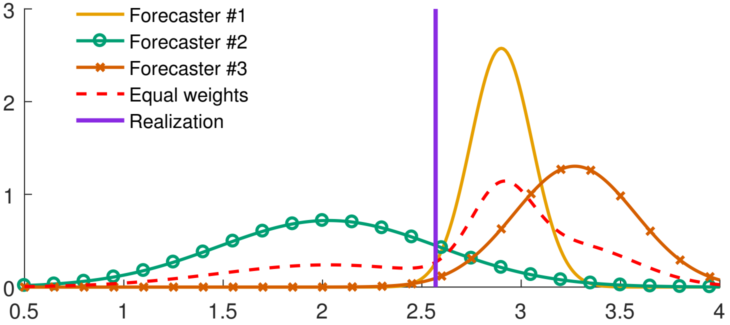
Note: Normal approximation based on midpoints of bins. Original bins: -3% to 6% with 1% increments.

1. Equal weights seem to perform well.

The problem of density forecast combinations

- Survey of Professional Forecasters: large-scale, quarterly macroeconomic survey maintained by the Philadelphia Fed.

Figure: Forecasters' predictions for US ΔGDP in 2014 as of 2014:Q1



Note: Normal approximation based on midpoints of bins. Original bins: -3% to 6% with 1% increments.

1. Equal weights seem to perform well.
2. Forecasts display patterns over time.

1. Combining densities: hedging against extremes.
2. Predictive distributions are complex objects, unlike point forecasts.
3. Equal weights often perform well, but ...
    - ...do not depend on data.

The problem of density forecast combinations: Summary

1. Combining densities: hedging against extremes.
2. Predictive distributions are complex objects, unlike point forecasts.
3. Equal weights often perform well, but ...
    - ...do not depend on data.
    - ...excellent forecasts receive just as much weight as implausible ones.

The problem of density forecast combinations: Summary

1. Combining densities: hedging against extremes.
2. Predictive distributions are complex objects, unlike point forecasts.
3. Equal weights often perform well, but ...
   - ...do not depend on data.
   - ...excellent forecasts receive just as much weight as implausible ones.
   - ...do not help understanding models' performance over time.
4. Why not exploit information on past forecasting performance?

The problem of density forecast combinations: Summary

1. Combining densities: hedging against extremes.
2. Predictive distributions are complex objects, unlike point forecasts.
3. Equal weights often perform well, but ...
   - ...do not depend on data.
   - ...excellent forecasts receive just as much weight as implausible ones.
   - ...do not help understanding models' performance over time.
4. Why not exploit information on past forecasting performance?

Can we do better?

The problem of density forecast combinations: Summary

1. Combining densities: hedging against extremes.
2. Predictive distributions are complex objects, unlike point forecasts.
3. Equal weights often perform well, but . . .
    - . . .do not depend on data.
    - . . .excellent forecasts receive just as much weight as implausible ones.
    - . . .do not help understanding models' performance over time.
4. Why not exploit information on past forecasting performance?

Can we do better?

YES!

I propose a formal, data-driven method to combine density forecasts, targeting the true conditional predictive density.

Contribution

1. Consistent estimator of weights that minimize discrepancy between combined forecast density and true predictive density (optimality) using goodness-of-fit statistics:
    - Kolmogorov–Smirnov, Cramer–von Mises, Anderson–Darling (Kullback–Leibler Information Criterion).

Contribution

1. Consistent estimator of weights that minimize discrepancy between combined forecast density and true predictive density (optimality) using goodness-of-fit statistics:
    - Kolmogorov–Smirnov, Cramer–von Mises, Anderson–Darling (Kullback–Leibler Information Criterion).
2. Monte Carlo simulations support the proposed methodology.
    - Recommendation for practitioners: use estimators based on Anderson–Darling statistic or the KLIC.

Contribution

1. Consistent estimator of weights that minimize discrepancy between combined forecast density and true predictive density (optimality) using goodness-of-fit statistics:
   - Kolmogorov–Smirnov, Cramer–von Mises, Anderson–Darling (Kullback–Leibler Information Criterion).
2. Monte Carlo simulations support the proposed methodology.
   - Recommendation for practitioners: use estimators based on Anderson–Darling statistic or the KLIC.
3. Empirical application: predicting US industrial production one month ahead using simple ARDL models.
   - Models' weights display time variation.
   - First study to find that spread and stock returns were valuable for *density* forecasts during the Great Recession (Ng and Wright (2013): point forecasts).
   - Housing data was of great help before and after the crisis.

Roadmap

1. Motivation
2. Econometric framework
   - Theory: building blocks
   - Theory: results
3. Monte Carlo results
4. Empirical application
   - Data and models
   - Results

Why and how to combine forecasts?

- Why combine density forecasts?
  1. Misspecification.
  2. Parameter estimation uncertainty, structural breaks.

Why and how to combine forecasts?

- Why combine density forecasts?
  1. Misspecification.
  2. Parameter estimation uncertainty, structural breaks.
- How to combine density forecasts?
  1. Large literature on density and point forecast evaluation (Diebold et al., 1998; Corradi and Swanson, 2006a,b,c; Rossi and Sekhposyan, 2013, 2014, 2016).
  2. Numerous theoretical and empirical results on optimal *point* forecast combinations (Bates and Granger, 1969; Stock and Watson, 2004; Cheng and Hansen, 2015).
  3. Few results on density forecast combinations (Hall and Mitchell, 2007; Geweke and Amisano, 2011; Kapetanios et al., 2015).

Why and how to combine forecasts?

- Why combine density forecasts?
  1. Misspecification.
  2. Parameter estimation uncertainty, structural breaks.
- How to combine density forecasts?
  1. Large literature on density and point forecast evaluation (Diebold et al., 1998; Corradi and Swanson, 2006a,b,c; Rossi and Sekhposyan, 2013, 2014, 2016).
  2. Numerous theoretical and empirical results on optimal *point* forecast combinations (Bates and Granger, 1969; Stock and Watson, 2004; Cheng and Hansen, 2015).
  3. Few results on density forecast combinations (Hall and Mitchell, 2007; Geweke and Amisano, 2011; Kapetanios et al., 2015).

Consistent estimator of density combination weights is of theoretical and practical importance.

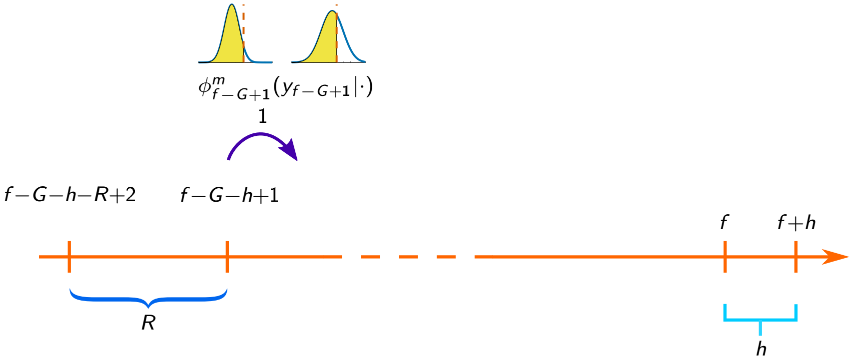
1. Motivation
2. Econometric framework
   - Theory: building blocks
   - Theory: results
3. Monte Carlo results
4. Empirical application
   - Data and models
   - Results

Notation I – Forecasting environment

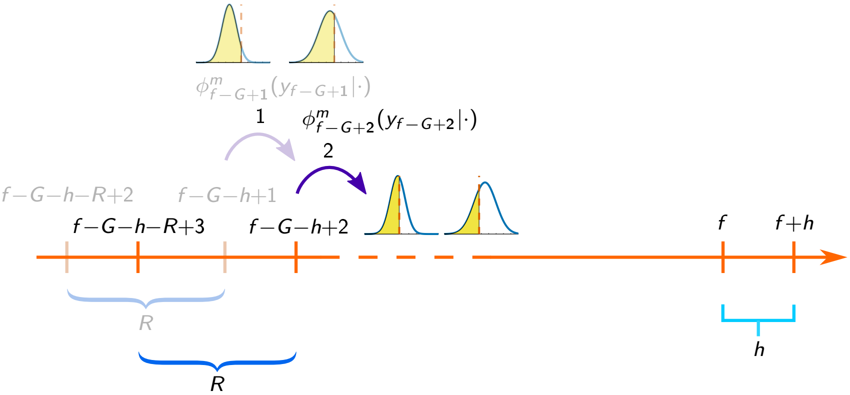
- y_{t+h} is the variable of interest and X_t is a vector of predictors.
- We want to forecast h periods ahead, $h < \infty$ and fixed.
- At forecast origin f, the researcher estimates models $m = 1, \ldots, \mathcal{M}$ in rolling windows of size R, where each estimation is based on the truncated information set $\mathcal{I}_{t-R+1}^{t}$, containing information between $t - R + 1$ and t.
- The total number of estimation windows is G.
- At each t, each model m implies a predictive density with typical element $\phi_{t+h}^{m}(y_{t+h}|\mathcal{I}_{t-R+1}^{t})$.
- $\phi_{t+h}^{*}(y_{t+h}|\mathcal{I}_{t-R+1}^{t})$ is the true conditional density.

Notation I – Illustration of estimation scheme

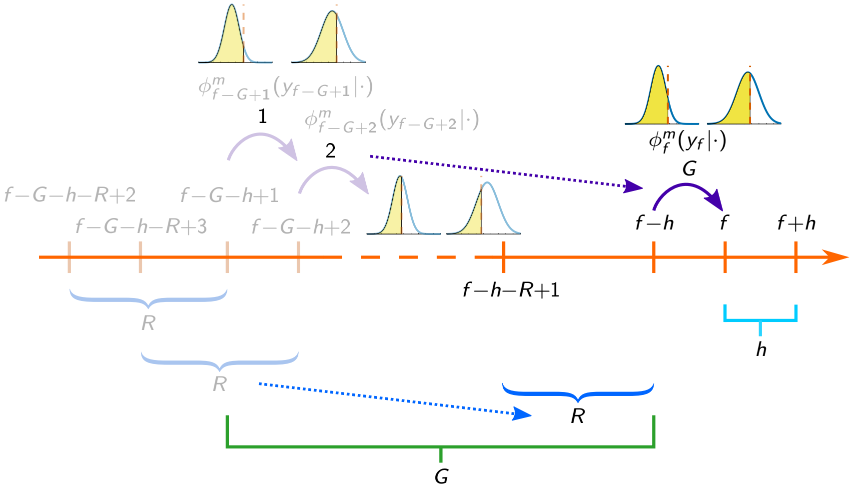
Notation I – Illustration of estimation scheme



Notation I – Illustration of estimation scheme



Notation I – Illustration of estimation scheme



Notation II – Combining density forecasts

- Researcher uses the convex combination of a set of $\mathcal{M}$ predictive densities:

$$\phi^{C}_{t+h}(y_{t+h}|\mathfrak{I}^{t}_{t-R+1}) \equiv \sum_{m=1}^{\mathcal{M}} w_m \phi^{m}_{t+h}(y_{t+h}|\mathfrak{I}^{t}_{t-R+1}). \qquad (1)$$

- Definition of probabilistic calibration: for a given w

$$\sum_{m=1}^{\mathcal{M}} w_m \phi^{m}_{t+h}(y_{t+h}|\mathfrak{I}^{t}_{t-R+1}) = \phi^{*}_{t+h}(y_{t+h}|\mathfrak{I}^{t}_{t-R+1}). \qquad (2)$$

Objective of this paper

Estimation of weights $w \in \Delta^{\mathcal{M}-1}$.

Notation III – Measuring (dis)similarity of distributions

- Probability integral transform or PIT is the combined CDF evaluated at the realization (Rosenblatt, 1952; Diebold et al., 1998):

$$\mathrm{PIT}_{t+h} \equiv \int_{-\infty}^{y_{t+h}} \phi^{C}_{t+h}(y|\mathcal{I}^{t}_{t-R+1})\,\mathrm{d}y = \Phi^{C}_{t+h}(y_{t+h}|\mathcal{I}^{t}_{t-R+1})\,.$$

- Kullback–Leibler Information Criterion (KLIC): the expected difference between true and combined log densities (White, 1994; Hall and Mitchell, 2007):

$$\mathrm{KLIC}(\Phi^{*}_{t+h}(y_{t+h}|\mathcal{I}^{t}_{t-R+1}), \Phi^{C}_{t+h}(y_{t+h}|\mathcal{I}^{t}_{t-R+1})) =$$

$$E_{\phi^*}\left\{\log \phi^{*}_{t+h}(y_{t+h}|\mathcal{I}^{t}_{t-R+1})\right\} - E_{\phi^*}\left\{\log \phi^{C}_{t+h}(y_{t+h}|\mathcal{I}^{t}_{t-R+1})\right\}\,.$$

  - The first term does not depend on w.

What makes a density forecast "good" and how to obtain it?

- Optimality: *given* the information (models) available at the forecast origin, we want to ensure that the combined forecast is probabilistically calibrated *or* as close to it as possible.
- Probabilistic calibration $\iff$ $\text{PIT}_{t+h} \sim \mathcal{U}(0,1)$ (Corradi and Swanson, 2006c; Gneiting et al., 2007). Does not assume knowledge of true DGP.
- Measures of closeness:
  - PIT-based: Kolmogorov–Smirnov, Cramer–von Mises and Anderson–Darling statistics.
  - Likelihood-based: KLIC.

Idea of this paper

Take a set of forecasting models and minimize the statistic of choice over combination weights $w \in \Delta^{\mathcal{M}-1}$.

Uniformity of the PIT I

- Optimistic forecaster:

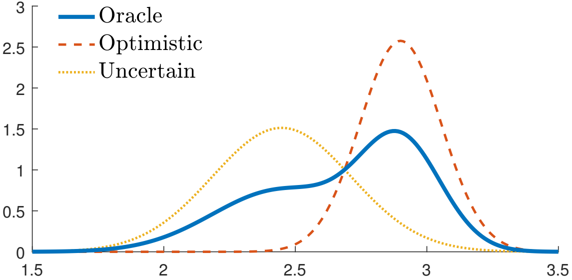
$$\phi^1_{t+1}(y_{t+1}|\mathcal{J}^t_{t-R+1}) = \mathcal{N}(2.90, 0.15^2)\,.$$

- Uncertain forecaster:

$$\phi^2_{t+1}(y_{t+1}|\mathcal{J}^t_{t-R+1}) = \mathcal{N}(2.45, 0.26^2)\,.$$

- "Oracle":

$$\phi^*_{t+1}(y_{t+1}|\mathcal{J}^t_{t-R+1}) = 0.5\mathcal{N}(2.90, 0.15^2) + 0.5\mathcal{N}(2.45, 0.26^2)\,.$$



Forecasters' predictive densities

Uniformity of the PIT II

Figure: CDFs of PITs

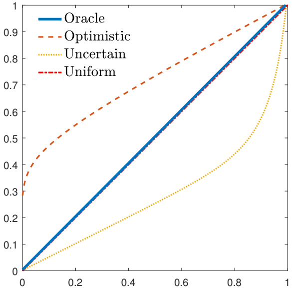
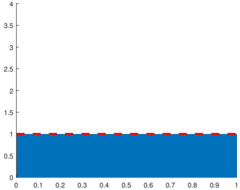
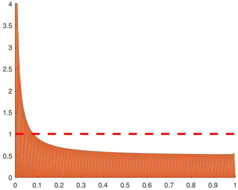


Figure: PDFs of PITs

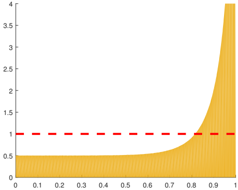
(a) "Oracle"



(b) Optimistic

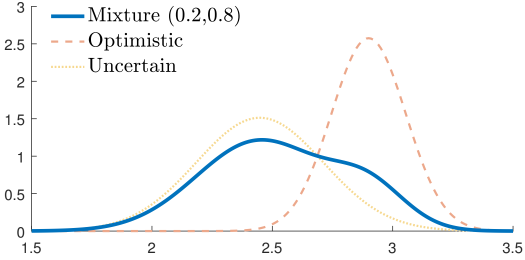


(c) Uncertain

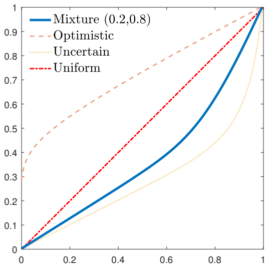


Uniformity of the PIT III - different weights

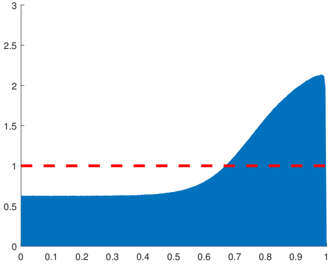
(a) Predictive densities



(b) CDF of PITs

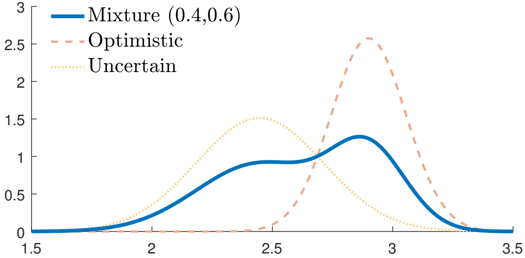


(c) PDF of mixture PIT

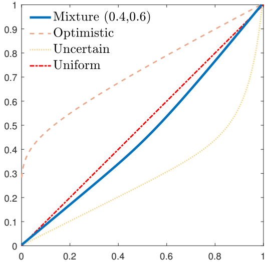


Uniformity of the PIT III - different weights

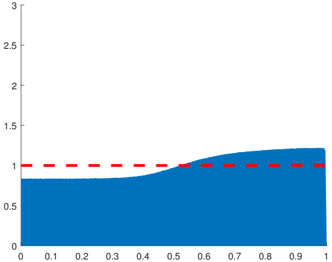
(a) Predictive densities



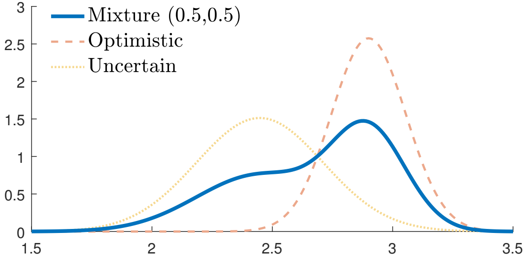
(b) CDF of PITs



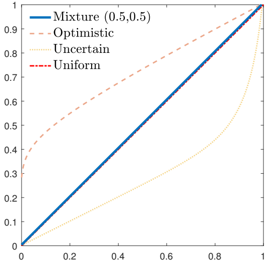
(c) PDF of mixture PIT



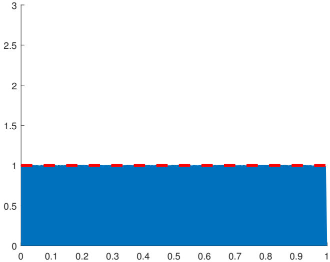
(a) Predictive densities



(b) CDF of PITs



(c) PDF of mixture PIT



Formal statistical problem I

- Vertical difference between CDF of PIT and uniform CDF (45° line) at quantile $r \in [0,1]$:

$$\Psi(r, w) \equiv P(\text{PIT}_{t+h} \leq r) - r \tag{3}$$

- Sample counterpart:

$$\Psi_G(r, w) \equiv G^{-1} \sum_{t=f-G-h+1}^{f-h} 1\left[\text{PIT}_{t+h} \leq r\right] - r \tag{4}$$

Formal statistical problem II

- Statistics used as objective functions, $T_G(w)$:

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

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

$$A_G(w) \equiv \int_0^1 \frac{\Psi_G^2(r, w)}{r(1-r)}\,\mathrm{d}r \qquad \text{(Anderson–Darling)}$$

$$\mathrm{KLIC}_G(w) \equiv -\frac{1}{G} \sum_{t=f-G-h+1}^{f-h} \log \phi_{t+h}^{C}(y_{t+h}|\mathcal{I}_{t-R+1}^{t}) \qquad \text{(KLIC)}$$

- Estimate w by minimizing the statistic of choice, formally:

$$\widehat{w} \equiv \underset{w}{\operatorname{argmin}}\, T_G(w) \tag{5}$$

Assumptions and consistency result

- M-estimators, proof builds on Newey and McFadden (1994), leading to strongly consistent estimators.
- Main assumptions:
  - $\{(y_t, X_t')'\}$ is weakly dependent (ϕ- or α-mixing).
  - Combined CDF $\Phi^C_{t+h}(y_{t+h}|\mathcal{I}^t_{t-R+1})$ is continuously distributed.
  - Combined pdf $\phi^C_{t+h}(y_{t+h}|\mathcal{I}^t_{t-R+1})$ is dominated (KLIC).
  - Rolling window estimation scheme: $R < \infty$ as $G, T \to \infty$, $1 \leq h < \infty$ and fixed.
  - Standard identification condition allowing for misspecification.

Theorem 1 (Consistency)

Under the assumptions stated above, the estimators are strongly consistent as $G \longrightarrow \infty$, that is $\widehat{w} \overset{a.s.}{\longrightarrow} w^*$, where w^* is the (pseudo) true weight vector.

1 Motivation

2 Econometric framework
- Theory: building blocks
- Theory: results

3 Monte Carlo results

4 Empirical application
- Data and models
- Results

Monte Carlo results in a nutshell

1. Consistency is clearly demonstrated.
2. Favorable results even for samples of size $G = 200$.
3. Estimator based on Anderson–Darling or KLIC statistic is recommended in practice (lowest MSE).
4. The paper contains a number of Monte Carlo simulations, all supporting the conclusions above:
    - AR processes with high/low persistence,
    - Bi- and trimodal densities,
    - Mixture of ARCH + AR models.

1. Motivation

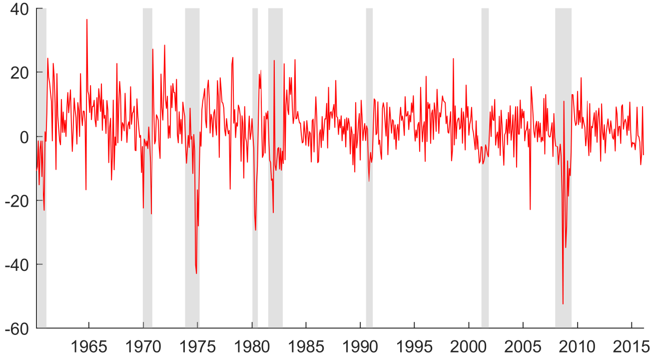
2. Econometric framework
   - Theory: building blocks
   - Theory: results

3. Monte Carlo results

4. Empirical application
   - Data and models
   - Results

Empirical application

Figure: Annualized US IP growth between March 1960 and February 2016



Note: Shaded areas are NBER recession periods.

- Predicting US industrial production (IP) growth one month ahead, using the Anderson–Darling objective function.

Modeling approach

Following Stock and Watson (2003), Granger and Jeon (2004), Gürkaynak et al. (2013) and Rossi and Sekhposyan (2014), I consider linear Autoregressive Distributed Lag (ARDL) models:

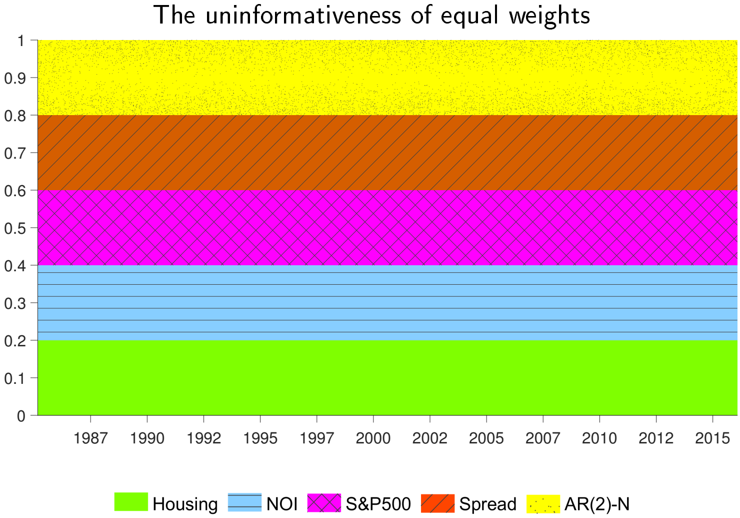
$$y_{t+1} = c + \sum_{j=0}^{1} \beta_j y_{t-j} + \sum_{j=0}^{1} \gamma_j x_{t-j} + \sqrt{\sigma^2}\varepsilon_{t+1},\ \varepsilon_{t+1} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$$

- y_t is annualized US IP growth, x_t is either New Private Housing Permits, ISM : New Orders Index, S&P 500, Moody's Baa Spread or $\emptyset$ (pure AR(2)). Data from May 2016 vintage of FRED-MD (McCracken and Ng, 2016).
- All five models estimated in rolling windows of $R = 120$ months, weights calculated over $G = 180$ months.
- Forecast target dates March 1985 – February 2016 ($P = 372$ months), which is the out-of-sample evaluation period.

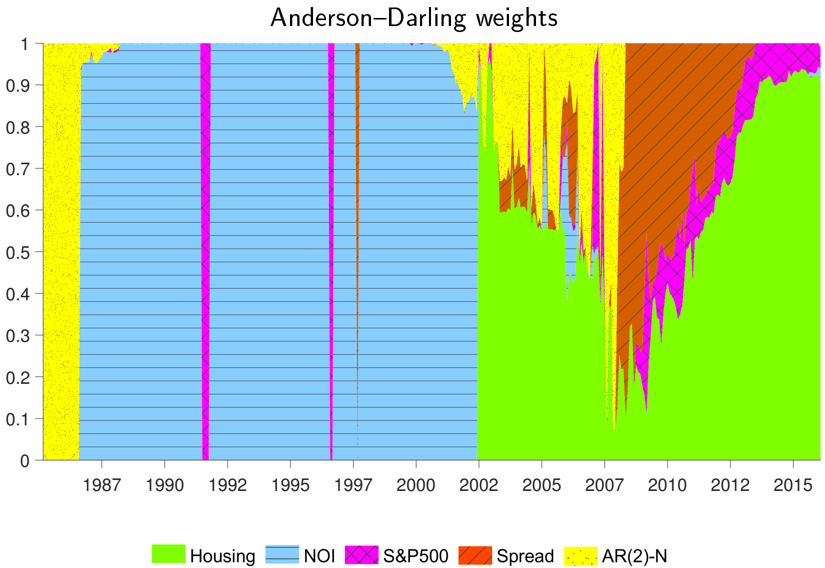
Benchmarks

1. KLIC (Hall and Mitchell, 2007; Geweke and Amisano, 2011).
2. AR(2)-N (Del Negro and Schorfheide, 2013).
3. Equal weights (Kascha and Ravazzolo, 2010; Rossi and Sekhposyan, 2014).
4. (Bayesian Information Criterion, Bayesian Model Averaging).

What do we learn from equal weights?



Time variation of estimated weights 1985:M3 – 2016:M2

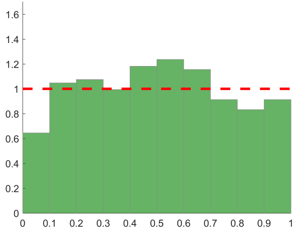


What did we learn about the models?

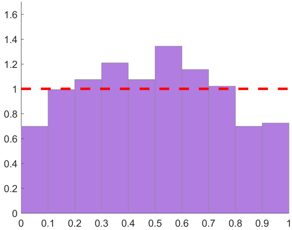
- No single model dominates the model set.
- Considerable time-variation of well-performing models.
- Anderson–Darling weights have economic meaning.
- Housing permits contains valuable predictive information before and after the Great Recession.
- S&P 500 and corporate spread receive large weights during the crisis. This is beyond the results of Ng and Wright (2013)!

Empirical results - PITs

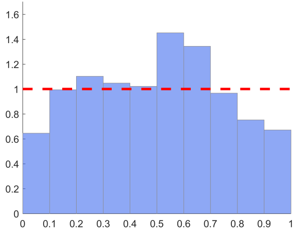
Figure: Normalized histograms of PITs



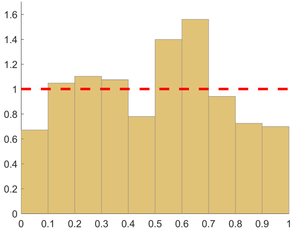
(a) AD weights



(b) KLIC weights



(c) Equal weights



(d) AR(2)-N model

Note: Horizontal red dashed line corresponds to uniform density.

The Anderson–Darling weights pass tests of uniformity!

Test statistics and p-values (in parentheses) of Rossi and Sekhposyan's (2016) test. H_0 : PIT is uniformly distributed.

Models	Kolmogorov–Smirnov	Cramer–von Mises
AD weights	0.90 (0.38)	0.24 (0.22)
KLIC weights	1.28 (0.08)	0.42 (0.06)
Equal weights	1.39 (0.05)	0.50 (0.04)
AR(2)-N	1.31 (0.08)	0.40 (0.09)

Note: p-values calculated with the HAC estimator by Newey and West (1987) using $\lfloor 0.75P^{1/3} \rfloor = 5$ lags. Number of Monte Carlo simulations was 200,000.

Conclusions and further research

1. Consistent weight estimator for density forecast combinations.
2. Theoretically appealing and works well in Monte Carlo simulations.
3. Empirical application the proposed framework delivers high quality density forecasts of US industrial production, outperforming the equal weights benchmark!
4. Weights are economically meaningful (housing bubble and leverage/spread).
5. First paper to find these patterns for *density* forecasts (Ng and Wright (2013): *point* forecasts only).
6. Potential extensions:
    - Penalized estimation of w, biasing to zero.
    - Testing for structural breaks.
    - Financial applications, such as Value-at-Risk.

The End

Thank you for your attention!

Additional results

- Knowledge of true DGP not required
- Monte Carlo DGP with estimated parameters
- Additional empirical results
- In-sample fit
- Robustness check

Example inspired by Corradi and Swanson (2006b,c)

- True DGP for y_{t+1} is a stationary normal AR(2) process.
- True predictive density of y_{t+1} conditional on $\mathcal{I}_t = \{y_t, y_{t-1}\}$:

$$\phi^*_{t+1}(y_{t+1}|\mathcal{I}_t) = \mathcal{N}(\alpha_1 y_t + \alpha_2 y_{t-1}, \sigma^2). \tag{6}$$

- The distribution of y_{t+1} conditional on y_t alone is also normal,

$$\phi^*_{t+1}(y_{t+1}|y_t) = \mathcal{N}(\widetilde{\alpha} y_t, \widetilde{\sigma}^2). \tag{7}$$

- If the researcher uses the AR(1) model instead of the AR(2) model, the forecast is still probabilistically calibrated, as given the information set (y_t), the predictive density is correct.

Probabilistic calibration does not require knowledge of the true DGP.

Back

Monte Carlo setup

- Constant and constant plus ARCH(1) model

$$M1 : y_{t+1} = c_1 + \nu_{t+1}$$

$$M2 : y_{t+1} = c_2 + \sqrt{\sigma^2_{2,t+1}}\varepsilon_{t+1}\,, \quad \sigma^2_{2,t+1} = \alpha_0 + \alpha_1 \varepsilon_t^2$$

$$\text{with } \nu_{t+1} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_1^2) \text{ and } \varepsilon_{t+1} \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$$

- DGP is a mixture of M1 and M2 with $(w_1, w_2)' = (0.4, 0.6)'$.
- Irrelevant M3 matches first 2 moments of true density:

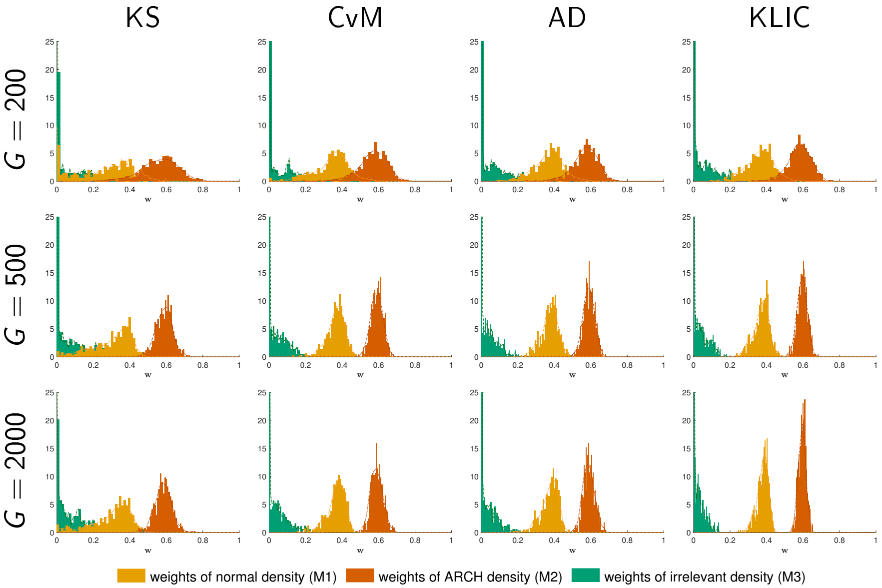
$$M3 : y_{t+1} = c_3 + \eta_{t+1} \qquad \eta_{t+1} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_3^2)$$

- $c_3 = w_1 \widehat{c}_1 + w_2 \widehat{c}_2$ and $\sigma_3^2 = w_1 \widehat{\sigma}_1^2 + w_2 \widehat{\sigma}^2_{2,t+1}$

Table: Parameters of M1, M2 and M3

Model	c	σ^2	α_0	α_1
M1	1	0.3	–	–
M2	1	–	0.2	0.2

Monte Carlo – Histograms, $w = (0.4, 0.6, 0)'$



Note: Histograms and kernel density estimates based on 2000 MC replications.

Back

Further benchmarks

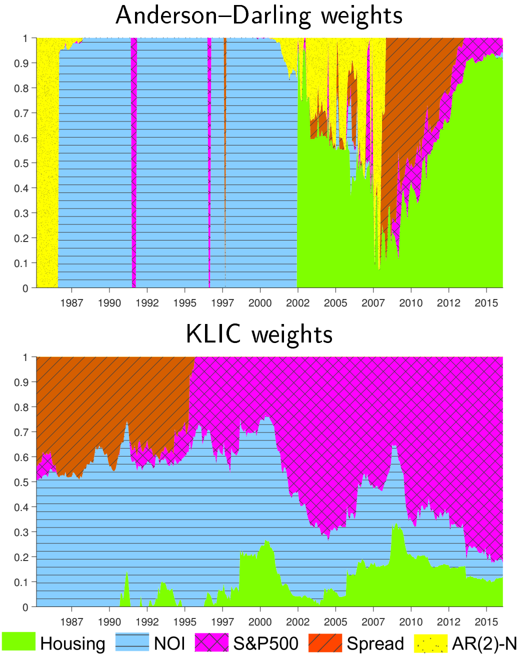
- Bayesian Information Criterion (Schwarz (1978); Kass and Raftery (1995); Hoeting et al. (1999)):

$$\text{BIC}_m \equiv -2 \sum_{t=f-R}^{f-1} \log l_m(y_{t+1}|z_t^m; \widehat{\theta}_m) + k_m \log(R) \,. \tag{8}$$

- Bayesian Model Averaging (Kass and Raftery (1995); Hoeting et al. (1999); Rossi and Sekhposyan (2014)):

$$w_m = \frac{\exp(-0.5BIC_m)}{\sum_{i=1}^{5} \exp(-0.5BIC_i)} \,. \tag{9}$$

Time variation of estimated weights 1985:M3 – 2016:M2



Time variation of estimated weights 1985:M3 – 2016:M2

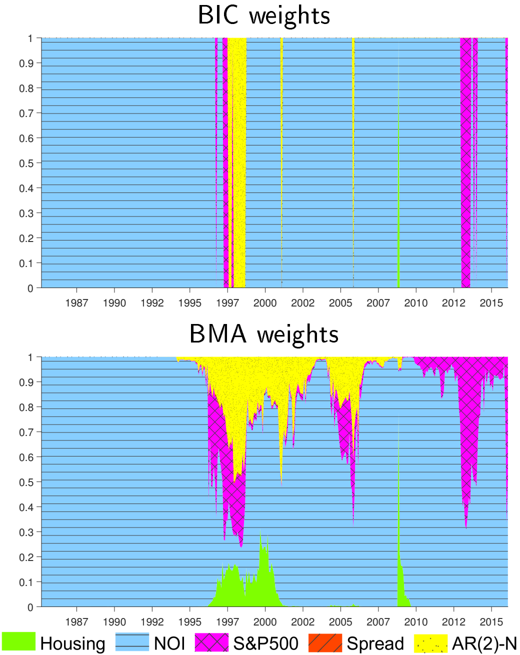
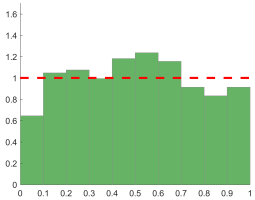
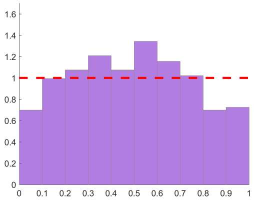


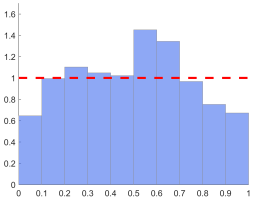
Figure: Normalized histograms of PITs



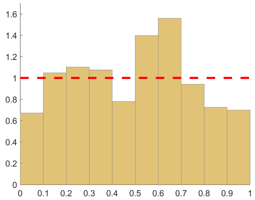
(a) AD weights



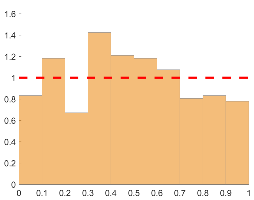
(b) KLIC weights



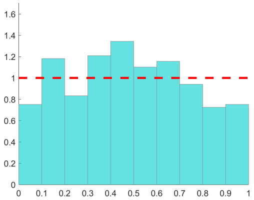
(c) Equal weights



(d) AR(2)-N model



(e) BIC weights



(f) BMA weights

Note: Horizontal red dashed line corresponds to uniform density.

Tests of uniformity of PITs

Rossi and Sekhposyan (2016) test on correct specification of conditional predictive densities. H_0 : PIT is uniformly distributed.

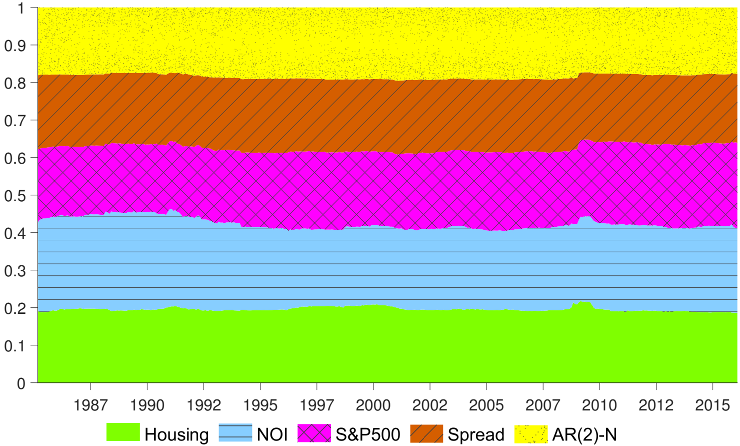
Models	Kolmogorov–Smirnov	Cramer–von Mises
AD weights	0.90 (0.38)	0.24 (0.22)
KLIC weights	1.28 (0.08)	0.42 (0.06)
Equal weights	1.39 (0.05)	0.50 (0.04)
AR(2)-N	1.31 (0.08)	0.40 (0.09)
BIC	1.16 (0.18)	0.32 (0.17)
BMA	1.28 (0.11)	0.38 (0.12)

Note: Test statistic (p-value). p-values calculated with the HAC estimator by Newey and West (1987) using $\lfloor 0.75P^{1/3} \rfloor = 5$ lags. Number of Monte Carlo simulations was 200,000.

Back

Did in-sample fit drive weights?

Figure: Ratios of inverse in-sample residual variances



Note: The sample period (end of the last rolling window of size $R = 120$) starts in February 1985 and ends in January 2016, with a total number of $P = 372$ months. Housing stands for Housing Permits, NOI is ISM: New Orders Index, while Spread is Moody's Baa Corporate Bond Yield minus Fed funds rate.

Back

Adding US house price index to the set of predictors

Back

References I

Bates, J. M. and Granger, C. W. J. (1969). The Combination of Forecasts. *OR*, 20(4):451–468.

Cheng, X. and Hansen, B. E. (2015). Forecasting with factor-augmented regression: A frequentist model averaging approach. *Journal of Econometrics*, 186(2):280–293.

Corradi, V. and Swanson, N. R. (2006a). Bootstrap conditional distribution tests in the presence of dynamic misspecification. *Journal of Econometrics*, 133(2):779–806.

Corradi, V. and Swanson, N. R. (2006b). Chapter 5 Predictive Density Evaluation. In Elliott, G., Granger, C. W. J., and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 1, pages 197–284. Elsevier.

References II

Corradi, V. and Swanson, N. R. (2006c). Predictive density and conditional confidence interval accuracy tests. *Journal of Econometrics*, 135(1-2):187–228.

Del Negro, M. and Schorfheide, F. (2013). DSGE Model-Based Forecasting. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2-A, pages 57 – 140. Elsevier, Amsterdam.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating density forecasts. *International Economic Review*, 39(4):863–883.

Geweke, J. and Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, 164(1):130–141.

References III

Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):243–268.

Granger, C. and Jeon, Y. (2004). Forecasting Performance of Information Criteria with Many Macro Series. *Journal of Applied Statistics*, 31(10):1227–1240.

Gürkaynak, R. S., Kisacikoglu, B., and Rossi, B. (2013). Do DSGE Models Forecast More Accurately Out-of-Sample than VAR Models? volume 32: VAR Models in Macroeconomics - New Developments and Applications: Essays in Honor of Christopher A. Sims of *Advances in Econometrics*, pages 27–79. Emerald Group Publishing Limited.

References IV

Hall, S. G. and Mitchell, J. (2007). Combining density forecasts. *International Journal of Forecasting*, 23(1):1–13.

Hoeting, J. A., Madigan, D. A., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science*, 14(4):382–417.

Kapetanios, G., Mitchell, J., Price, S., and Fawcett, N. (2015). Generalised density forecast combinations. *Journal of Econometrics*, 188(1):150–165.

Kascha, C. and Ravazzolo, F. (2010). Combining inflation density forecasts. *Journal of Forecasting*, 29(1-2):231–250.

Kass, R. E. and Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90(430):773–795.

References V

McCracken, M. W. and Ng, S. (2016). FRED-MD: A Monthly Database for Macroeconomic Research. *Journal of Business & Economic Statistics*, 34(4):574–589.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In McFadden, D. and Engle, R., editors, *Handbook of Econometrics*, volume 4, pages 2111–2245. Elsevier, Amsterdam.

Newey, W. K. and West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3):703.

Ng, S. and Wright, J. H. (2013). Facts and Challenges from the Great Recession for Forecasting and Macroeconomic Modeling. *Journal of Economic Literature*, 51(4):1120–1154.

References VI

Rosenblatt, M. (1952). Remarks on a Multivariate Transformation. *Ann. Math. Statist.*, 23(3):470–472.

Rossi, B. and Sekhposyan, T. (2013). Conditional predictive density evaluation in the presence of instabilities. *Journal of Econometrics*, 177(2):199–212.

Rossi, B. and Sekhposyan, T. (2014). Evaluating predictive densities of US output growth and inflation in a large macroeconomic data set. *International Journal of Forecasting*, 30(3):662–682.

Rossi, B. and Sekhposyan, T. (2016). Alternative Tests for Correct Specification of Conditional Predictive Densities. Working Paper No. 758, Barcelona GSE.

References VII

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464.

Stock, J. H. and Watson, M. W. (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, 41(3):788–829.

Stock, J. H. and Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23(6):405–430.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Number 22 in Econometric Society Monographs. Cambridge University Press, Cambridge.