Forecast Densities for Economic Aggregates from Disaggregate Ensembles

Francesco Ravazzolo[a]
Shaun Vahey[b]

[a]Norges Bank [b]ANU

May 6, 2011

Index construction

- Suppose that a set of N disaggregate prices $d = 1, .., N$ is simulated as:

$$y_{t,d} = \mu_{t,d} + \rho_{t,d} y_{t,d} + \epsilon_{t,d}, \; t = 1, ..., T, \; \epsilon_{t,d} \sim N(0, \sigma^2_{t,d}). \tag{1}$$

  where $y_{t,d} = (p_{t,d} - p_{t-1,d})$.

- Then, define an aggregate index as:

$$y_t = \sum_{d=1}^{N} w_{t,d} y_{t,d}, \; w_{t,d} > 0, \; \sum_{d} w_{t,d} = 1. \tag{2}$$

- Example: inflation, GDP,...

Model uncertainty issues

- Index weights (d, N) could be unknown and change next period (Lutkepohl, 2009 and 2010).
- Disaggregate i could be related to past values of disaggregate j (van Garderen *et al*, 2000).
- Time variation in the weights and model parameters introduces nonlinearities (Groen *et al.*, 2009).
- Data and weights could contain non trivial errors (Ravazzolo and Vahey, 2009, on Australia inflation).
- Aggregate might not have Gaussian errors (Cheung and Chung (2009) for long memory and Garch effects).
- **How to forecast future inflation?**

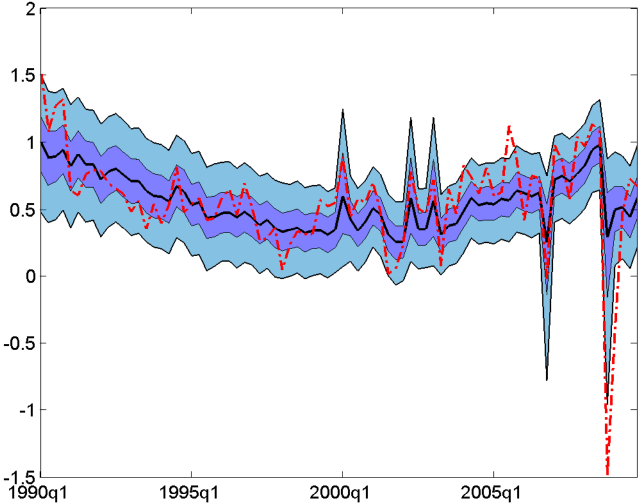
Ensemble modeling

- Uncertainty about model specifications (e.g. initial conditions, data, parameters, and boundary conditions) in dynamic, stochastic or nonlinear (chaotic) systems.
- Approximate (future) unknown non-linear non-Gaussian system with combination of (density) forecasts from large number of locally-linear models with time-varying weights.
- Use forecast density combination technology (Wallis, 2010; Hall and Mitchell, 2005; Jore *et al.*, 2010; Kascha and Ravazzolo, 2010) to construct ensemble predictive densities.
- The true model is not be included in the model space (Geweke, 2009).
- **Results:** our ensemble outperforms aggregate benchmarks in density forecasting both in simulation exercises and in an empirical application with US PCE inflation.

Outline

- Density forecasting
- Ensemble methodology
- Application
- Results
- Conclusions
- NB policy tool

Density forecasts (fanchart)



Evaluating Predictive Densities

Problem: The true density is never observed.

There are two main approaches in the literature:

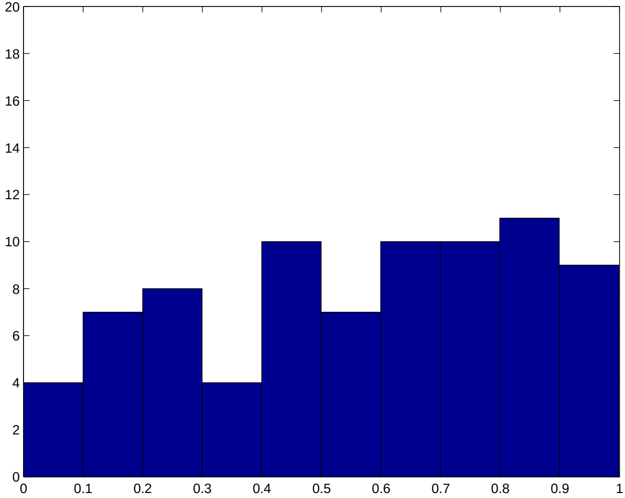
- Test whether a predictive density is correctly specified using Probability Integrated Transform Scores.
- Evaluate multiple, possibly misspecified and nested models:
  - Economic value of forecasts (finance).
  - With respect to a benchmark density.
  - Kullback-Leibler Information Criterion.
  - Continuous Ranked Probability Scores

PITS

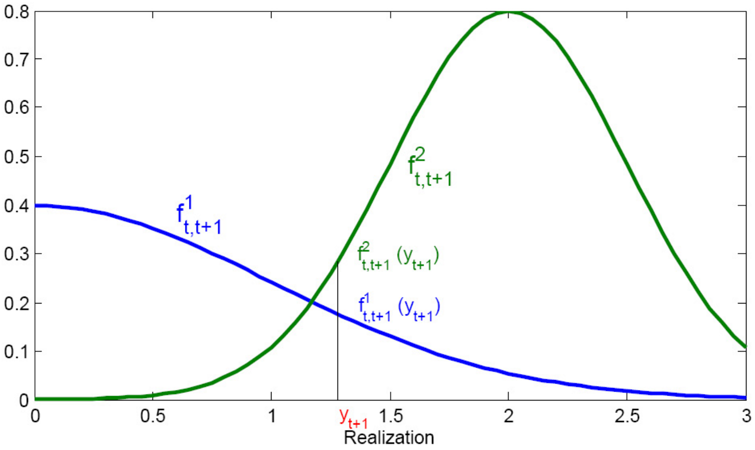
A forecast density is preferred if the density is correctly calibrated, regardless of the forecasters loss function:
The *PITS* are:

$$z_\tau = \int_{-\infty}^{\pi_\tau} g(u)du.$$

The *PITS* should be both uniformly distributed, and independently and identically distributed if the forecast densities are correctly calibrated. Hence, calibration evaluation requires the application of tests for goodness-of-fit and independence.


20
18
16
14
12
10
8
6
4
2
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

KLIC



0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.5
1
1.5
2
2.5
3
$f^2_{t,t+1}$
$f^1_{t,t+1}$
$f^2_{t,t+1}(y_{t+1})$
$f^1_{t,t+1}(y_{t+1})$
y_{t+1}
Realization

The KLIC distance between the true density $f(\pi_\tau)$ of π_τ and the ith density forecast, $f(\pi_\tau \mid I_{i,\tau})$, for π_τ conditional on information up to time $\tau - 1$ is defined as

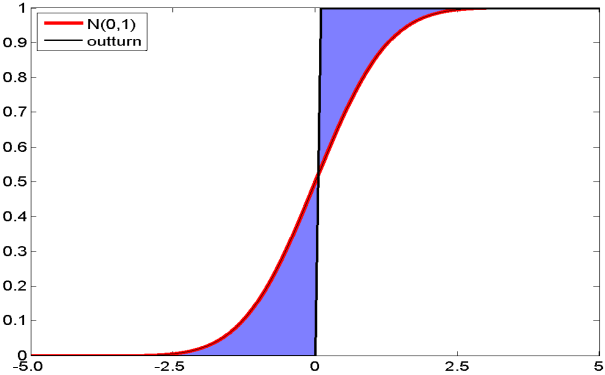
$$\begin{aligned} \mathrm{KLIC}_\tau &= \int f(\pi_\tau) \ln \frac{f(\pi_\tau)}{f(\pi_\tau \mid I_{i,\tau})} d\pi_\tau \\ &= E[\ln f(\pi_\tau) - \ln(\pi_\tau \mid I_{i,\tau})]. \end{aligned}$$

The latter term on the right hand side can be estimated by the average logarithmic score or "log-score"

$$\ln S^i := -\frac{1}{T^e} \sum_{t=T^s}^{T} \ln(\pi_\tau \mid I_{i,\tau}),$$

$T^e = T - T^s + 1.$

CRPS



- The CRPS is measured as the difference between the predicted and actual cumulative distribution. The (positive) score approaches zero as the predictive density converges on the true (but unobserved) density.
- In formula:

$$CRPS = \int_{-\inf}^{+\inf} (F(h) - I(h \geq \pi_o))^2 dh \tag{3}$$

where $I(A) = 1$ if A is true.
- Approximated (Panagiotelis and Smith 2008) by:

$$CRPS = E_F|\pi - \pi_o| - 0.5E_F|\pi - \pi'| \tag{4}$$

where E_F is the expectation for the predictive F, π and π' are independent random draws from the predictive, and π_o is the observed outturn.

Ensembles

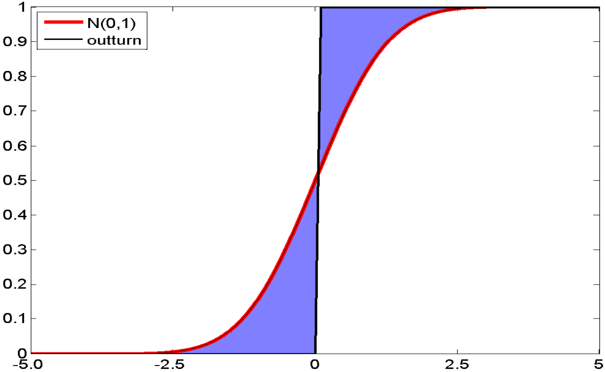
- Given $i = 1, \ldots, N$ different model specifications:

$$p(\pi_\tau) = \sum_{i=1}^{N} w_{i,\tau} \; g(\pi_\tau \mid I_{i,\tau}), \qquad \tau = \underline{\tau}, \ldots, \overline{\tau}. \qquad (5)$$

- Where $g(\pi_\tau \mid I_{i,\tau})$ is the **density** of interest from component i, $i = 1, \ldots, N$.
- And $w_{i,\tau} \geq 0$, $\sum_i w_{i,\tau} = 1$;
$w_{i,\tau} = \frac{\left[\sum_{\underline{\tau}}^{\tau-1} X(g(\pi_\tau | I_{i,\tau}))\right]}{\sum_{i=1}^{N}\left[\sum_{\underline{\tau}}^{\tau-1} X(g(\pi_\tau | I_{i,\tau}))\right]}, \qquad \tau = \underline{\tau}, \ldots, \overline{\tau}.$

- X is a CRPS based measure (Continuous Ranked Probability Scores).
- Generation of combined densities with component model weights varying through time.

CRPS



- Inflation predictive densities for measured inflation based on disaggregate inflation series (**uncertainty on model specifications**).
- Construct predictive densities for measured inflation from [AR(2)] (dis)aggregates $g(\pi_\tau \mid I_{i,\tau}),\ i = 1, \ldots, N$ using moving window estimation (**locally-linear**).
- Ensemble these predictive densities via linear opinion pool (**non-linear non-Gaussian system**) to forecast 1-step ahead headline PCE deflator inflation:
  $p(\pi_\tau) = \sum_{i=1}^{N} w_{i,\tau}\ g(\pi_\tau \mid I_{i,\tau})$.
- Correct the bias to produce predictive densities for measured inflation using a $\tau - \underline{\tau} = 20$ quarter moving window.

Simulation

- Two disaggregates $d = 1, 2$ are simulated as:

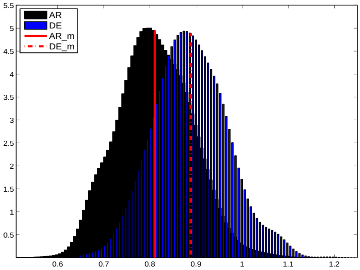
$$y_{t,d} = \mu_{t,d} + \rho_{t,d} y_{t,d} + \epsilon_{t,d}, \; t = 1, ..., T, \; \epsilon_{t,d} \sim N(0, \sigma^2_{t,d}). \tag{6}$$

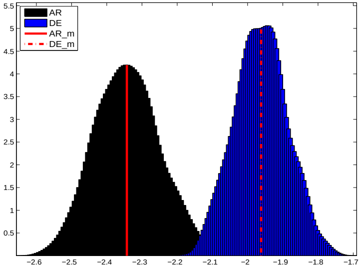
- The aggregate is then computed as:

$$y_t = \sum_d w_{t,d} y_{t,d}, \; w_{t,d} > 0, \; \sum_d w_{t,d} = 1. \tag{7}$$

- T =120; initial sample period=40; training period= 20 forecasts; OOS=60 observations. Simulation=1000 times.
- **Time varying means, standard deviations and weights:** two breaks at time $t = 20$ and $t = 60$ on $\mu_{20,d}$ and $\mu_{60,d}$, and $\sigma_{20,d}$ and $\sigma_{60,d}$, time-varying [AR] weights.
- AR(1) benchmark for aggregate.
- Time varying weights and parameters, non-linear non-Gaussian system, true model unknown.
- Two disaggregates, independent disaggregates.

RMSPE - LS





US Data

Quarterly Personal consumption expenditures (PCE) inflation (and disaggregate series) over the sample 1975Q1-2009Q4.

Group 1:

16 disaggregates: Motor vehicles and parts - Furnishings and durable household equipment - Recreational goods and vehicles, Other durable goods - Food and beverages - Clothing and footwear - Gasoline and other energy goods - Other nondurable goods - Housing and utilities - Health care - Transportation services - Recreation services - Food services and accommodations - Financial services and insurance - Other services - Final consumption expenditures of nonprofit institutions serving households.

Application: US PCE inflation forecasting

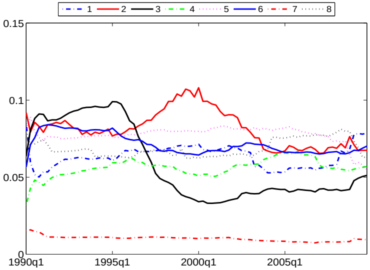
- Out-of-sample period: 1990Q1-2009Q4.
- Training period: 1985Q1-1989Q4.
- Estimation: moving window of 40 observations.
- Benchmarks: AR(2) and IMA for PCE.
- Disaggregate ensemble: DE11 based on AR(2) models for the 16 disaggregates.
- Geweke and Amisano (2009) and Mitchell and Wallis (2010) apply absolute and relative measures.
- *Absolute:* Test whether a predictive density is correctly specified (Diebold, Gunther and Tay, 1998; **Berkowitz 2001**, **Anderson-Darling**, **Pearson chi-squared**, **Ljung-Box tests**).
- *Relative:* based on the KLIC distance and rewards models which on average give higher probability to events which have actually occurred (**Mitchell and Hall, 2005** and **Amisano and Giacomini, 2007**).

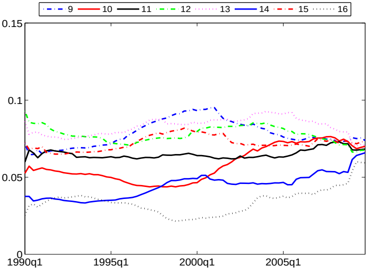
Density forecast performance

	LR3	AD	χ^2	LB	LS	LS_test
DE	**0.051**	**0.097**	**0.408**	**0.520**	**0.064**	**0.000**
			Individual models			
AR2	0.000	0.000	0.000	**0.504**	-0.465	
IMA	0.014	0.000	0.000	**0.871**	-0.095	**0.000**

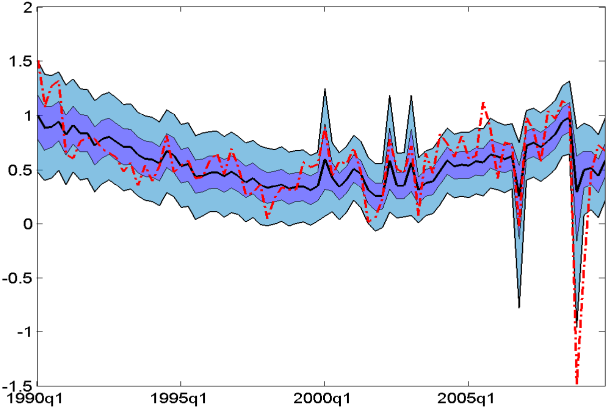
LR3=Berkowitz (2001), AD=Anderson-Darling, χ^2=Pearson chi-squared, LB=Ljung-Box; LS=log scores.

DE11's weights





DE11 interval forecasts



Conclusions

- John Kay (FT, September 21 2010): "There will always be a demand for forecasts, so there will always be a supply. But the reputation of economic forecasters, like other quacks and charlatans, depends more on the slickness of their presentations than the value of their work".
- Does he mean **POINT** forecasts?
- Density forecasting is a necessary input into economic decision making (Granger and Pesaran, 2000).

Conclusions

- Conventional view: **POINT** forecasting economic aggregate with disaggregates does not work, although used routinely by monetary policy makers.

- Our focus is on **DENSITY** forecasting with model uncertainty and incomplete component model space.
- Density ensemble strategy works well in Monte Carlo and in PCE application.

- Further applications:
  - Larger number of inflation disaggregates and international comparison.
  - Euro-area GDP.
  - Forecasting disaggregates with disaggregates.

Norges Bank core inflation measure

- NB applies the methodology to a relative large set of disaggregate CPI inflation (from 40 to 45 disaggregates) to construct a core inflation measure.
- NB forecasts what knows, discarding noise. This can be interpreted as a definition of core inflation measure.
- It allows to understand whether a shock to a specific component has a pass throw mechanism to the inflation process.

Norges Bamk MP2010/3