Conference on

Forecasting and Monetary Policy

Berlin, 23-24 March 2009

James Mitchell
NIESR London

„Combining Forecast Densities from VARs and DSGEs with Uncertain Instabilities“

niesr

Combining VAR and DSGE Forecast Densities

Ida Wolden Bache *(Norges Bank)*
Anne Sofie Jore *(Norges Bank)*
James Mitchell *(NIESR, London)*
Shaun Vahey *(Melbourne Business School)*

National Institute of Economic and Social Research

Density forecasting

- Density forecasts or “fan charts” provide an estimate of the probability distribution of a variable’s possible future values
- They provide a full impression of the uncertainty associated with a forecast
  - Point forecasts are better seen as the central points of ranges of uncertainty
  - It is not a question of this (point) forecast proving to be right and that forecast proving to be wrong
  - It can be rational to produce “biased” point forecasts with an asymmetric loss function (like many central banks?)

niesr

Ensemble density forecasting

- A popular macroeconomic forecasting strategy takes combinations across many models to hedge against model instabilities of unknown timing
  - Stock & Watson (2004) and Clark & McCracken (2009) find that equal-weighting of component models produces good point forecasts. Re-assuring for those with quadratic loss functions
  - Jore, Mitchell & Vahey (2009) find that VAR components weighted by the logarithmic score produce well-calibrated ensemble densities for real-time US data
- But they exclude DSGE models
- This is an obstacle to the implementation of ensemble forecasting methods at central banks

niesr

DSGE and VAR components

- We use an expert combination framework to combine forecast densities from VARs and a DSGE model (NEMO - the Norges Bank core policymaking model)
- Component forecasts are combined to produce the ensemble forecast density using the logarithmic score
- We evaluate the forecast densities using probability integral transforms (*pits*). This offers a means of evaluating density forecasts for general but unknown loss functions

niesr

Ensemble Forecasts

- Following Morris (1974, 1977) and Winkler (1981) we adopt an “expert” combination approach
- We define $i=1,\ldots,N$ experts, where each expert produces one of the N density forecasts
- A decision maker (*DM*) combines the densities from two or more experts based on the fit of the density forecasts over the Evaluation Period (*EP*)
  - The experts do not interact prior to combination by the *DM*
- The *DM* learns about the predictive densities by combining prior information with the evidence presented by the experts in the *EP* using Bayes’ rule

Linear opinion pool

- The *DM* constructs the combined densities by a linear opinion pool method (Wallis 2005; Mitchell & Hall, 2005)

$$p(Y_{\tau,h}) = \sum_{i=1}^{N} w_{i,\tau,h} g(Y_{\tau,h} \mid I_{i,\tau})$$

  where $g(Y_{\tau,h} \mid I_{i,\tau})$ are the h-step ahead forecast densities of expert *i* of realisation y_τ using information set $I_{i,\tau}$ which given publication delays includes information dated $\tau - 1$ and earlier
- $w_{i,\tau,h}$ are a set of non-negative weights that sum to unity
- The linear opinion pool delivers a more flexible distribution than each of the individual densities and accounts for model uncertainty

niesr

Choosing the weights

- Recursive logarithmic score weights (RLSW)
  - We use the log score (*logS*) to measure the fit of the experts' densities through the *EP*

$$w_{i,\tau,h} = \frac{\exp\left[\sum_{\underline{\tau}-10}^{\tau-1-h} \ln g(y_{\tau,h} \mid I_{i,\tau})\right]}{\sum_{i=1}^{N} \exp\left[\sum_{\underline{\tau}-10}^{\tau-1-h} \ln g(y_{\tau,h} \mid I_{i,\tau})\right]}; \quad \tau = \underline{\tau}, \ldots, \overline{\tau}$$

  - From a Bayesian perspective this approach has many similarities with an approximate predictive likelihood approach (BMA)
    - Given our definition of density fit, the densities are combined using Bayes' rule with equal (prior) weight on each model – which a Bayesian would term non-informative priors

niesr

Forecast density evaluations

- Density forecasts are evaluated by testing whether the probability integral transforms of the forecast density with respect to the realisation of the variable are uniform and i.i.d. (for *h=1*)
  - Goodness-of-fit tests
    - Consider the LR test proposed by Berkowitz (2001) with for *h>1* a 2 degrees-of-freedom variant; the Anderson-Darling test; a Pearson chi-squared test (Wallis, 2003)
  - Independence tests. Use a Ljung Box test
  - Since more than one density forecast can produce uniform even independent *pits*, as suggested by Mitchell & Wallis (2009), we also consider KLIC-based tests based on differences in logarithmic scores (*logS*)

niesr

Component models

- VAR
  - Itself an ensemble, constructed by combining many component VAR densities using the log score weights and a 10 observation training period
  - Construct variants both with constant parameters and allowing for breaks
    - these are the “no break VAR ensemble” and the “break VAR ensemble”
- DSGE

niesr

Many VAR component models

- Consider a range of 30 VARs in inflation, output growth and the interest rate
  - Various sizes of VAR (incl. ARs), lag lengths, transformations of variables (first differences, de-trended)
- *h*-step ahead forecasts are computed via the *direct* approach for a given (constant parameter) VAR model:

$$Y_t = \alpha + \beta X_{t-h} + \sigma\varepsilon_t ;\ \varepsilon_t \sim N(0,1)$$

- The predictive densities for Y_{t+h} (with non-informative priors) are multivariate Student-*t* (see Zellner, 1971)

niesr

VAR components – with breaks

- Following Garratt, Koop & Vahey (2008, EJ) and Jore, Mitchell & Vahey (2009) we deal with structural breaks, in the conditional mean and/or variance, by estimating a given VAR model, for a given recursion, with each candidate break date
- This is a convenient way to tackle break-date uncertainty

VAR models – with breaks
Computational details

- For computational simplicity
  - we restrict the break dates to be identical across equations for each VAR model
  - consider every feasible break date value with a regime containing at least 40% of the observations
  - restrict the maximum number of regimes, R, to 3
- With these additional structural break models added to the set of full sample VARs, we consider a maximum of 810 component models (for the final recursion in our EP)
  - We also estimate the 30 models over a rolling window to give 840 component models in total. Rolling regression models are advocated as a means of tackling structural breaks by (among others) Eklund, Kapetanios & Price (2008)

niesr

DSGE model

- NEMO (the Norwegian Economy MOdel) is the core model used for structural analysis at Norges Bank
- It is a medium-scale New Keynesian small open economy model with a similar structure to the DSGE models recently developed in many central banks
- We estimate the structural parameters using Bayesian techniques
  - Estimation uses data on 11 variables, including those in the VAR

niesr

The Norwegian data

- Our recursive forecasting exercise is intended to mimic the behaviour of a policymaker forecasting in real time
  - Unfortunately, we are not able to utilise real-time macroeconomic data because the data have not been compiled for (most of) the 11 observables used in DSGE estimation
- We use a single vintage of data available in 2008q4 for all forecasts and realisations
- Evaluation Period is 1997q3 to 2007q3 (estimation sample starts in 1987q1 to match Norges Bank)
- Focus on forecasts of inflation

niesr

Results: RMSFE

Horizon	AR(1)	DSGE	break VAR ensemble	no break VAR ensemble
$h = 1$	0.80	0.86	0.77	0.77
$h = 2$	0.81	1.01	0.82	0.83
$h = 3$	0.95	1.07	0.84	0.88
$h = 4$	1.01	1.23	0.94	1.02

- Consistent with SW and CM etc., VAR ensembles (point) forecast effectively
- The DSGE is not as effective as these ensembles
- The DSGE is competitive against a single constant-parameter VAR (the traditional benchmark)
- At longer horizons the break ensemble does better

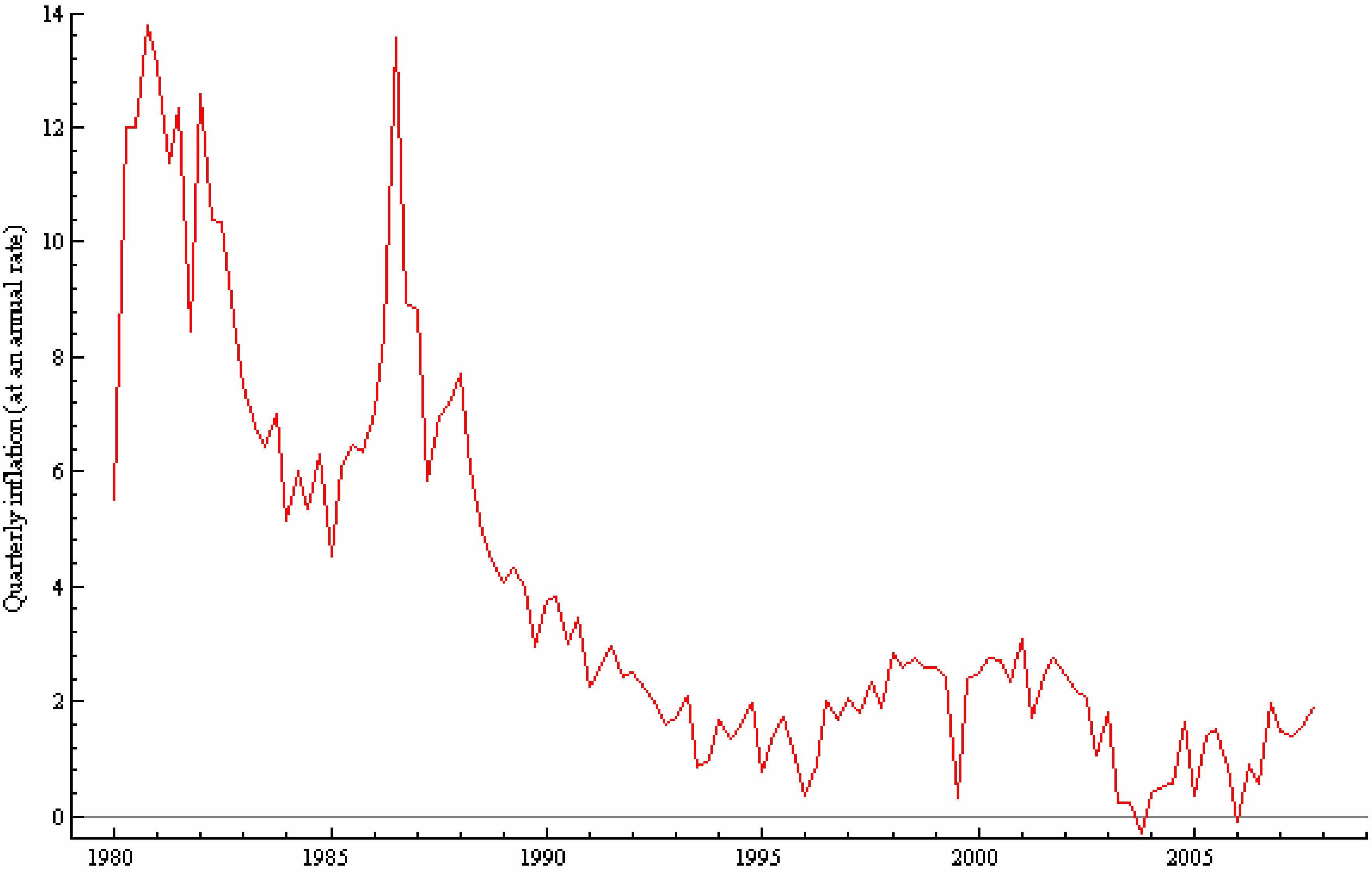
niesr

Results: component weights

- Since we focus on performance of the ensemble, rather than its components, we do not report the individual model weights (RLSW) for the many structural break variants in the break VAR ensemble
- But we note that the break models with the most support typically have a single break in the interval 1985 to 1987

niesr

Norwegian inflation



Density forecast evaluation using the *pits*

$h = 1$	LR3	AD	χ^2	LB1	logS
no break VAR ensemble	0.00	1.80	0.01	0.56	-1.287
break VAR ensemble	0.43	0.96	0.13	0.83	-1.185
DSGE	0.00	3.27	0.02	0.02	-1.265

$h = 2$	LR2	AD	χ^2	MLB	logS
no break VAR ensemble	0.00	2.61	0.01	0.86	-1.390
break VAR ensemble	0.37	1.16	0.32	0.94	-1.223
DSGE	0.00	7.11	0.00	0.82	-1.440*

$h = 3$	LR2	AD	χ^2	MLB	logS
no break VAR ensemble	0.08	0.83	0.20	0.75	-1.388
break VAR ensemble	0.36	1.07	0.49	0.85	-1.319
DSGE	0.00	9.36	0.00	0.91	-1.590*

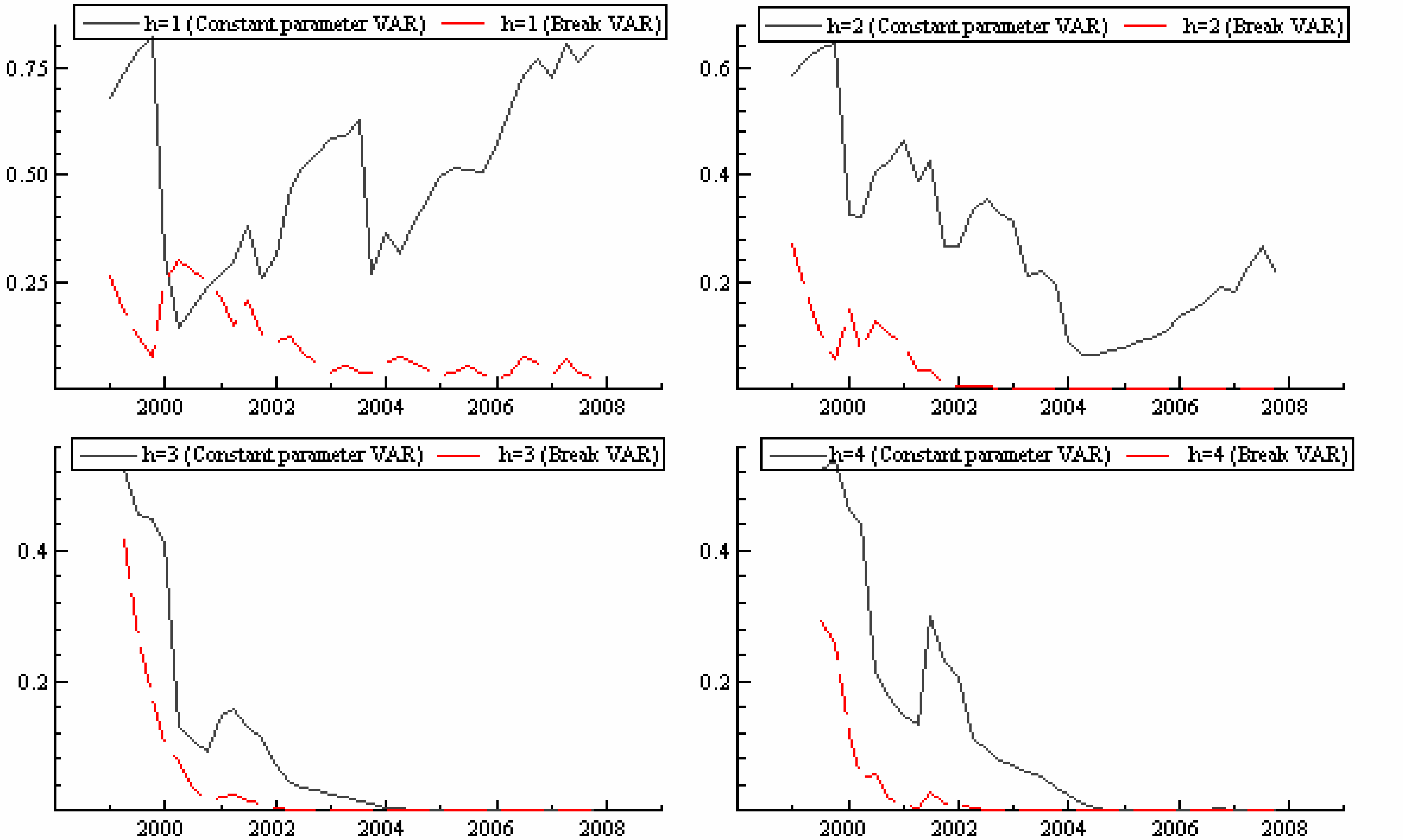
$h = 4$	LR2	AD	χ^2	MLB	logS
no break VAR	0.10	0.77	0.87	0.68	-1.537
break VAR	0.06	2.36	0.12	1.00	-1.487
DSGE	0.00	9.15	0.00	0.93	-1.706

The grand ensemble

- Unlike the others, the break VAR ensemble produces well-calibrated density forecasts, since it gives substantial weight to models that allow for structural breaks in the mid 1980s
- It is also instructive to combine the VAR and DSGE ensemble densities using the recursive log scores
  - Garratt, Mitchell & Vahey (2009) refer to a combination of ensemble predictive densities as a “grand ensemble”
  - Here one of the candidates is not an ensemble--the DSGE predictive density--but the methodology is similar

niesr

Recursive logarithmic score weights on the DSGE in the grand ensemble



Density forecast evaluation of the grand ensemble using the *pits*

$h=1$	LR3	AD	χ^2	LB1	logS
grand ensemble I	0.01	**2.08**	**0.42**	**0.48**	-1.270
grand ensemble II	**0.33**	**1.16**	**0.06**	**0.65**	-1.193

$h=2$	LR2	AD	χ^2	MLB	logS
grand ensemble I	0.00	3.43	0.01	**0.91**	-1.411*
grand ensemble II	**0.32**	**1.32**	**0.40**	**0.91**	-1.226

$h=3$	LR2	AD	χ^2	MLB	logS
grand ensemble I	**0.08**	**0.94**	**0.08**	**0.78**	-1.413
grand ensemble II	**0.31**	**1.19**	**0.60**	**0.81**	-1.332

$h=4$	LR2	AD	χ^2	MLB	logS
grand ensemble I	**0.05**	**1.21**	**0.82**	**0.70**	-1.561
grand ensemble II	**0.07**	**2.39**	**0.12**	**1.00**	-1.487

Conclusion

- Ensemble densities based on component VARs with breaks are well calibrated
- Ensembles from constant parameter VARs are typically poorly calibrated
- Although many policymakers prefer DSGE models for structural analysis, the forecast densities do not match the performance of break VAR ensembles
- The DSGE model receives a low weight in the grand ensemble with break VARs
- When combined with constant parameter VARs, the densities from the DSGE receive a substantial weight at some horizons

niesr

The way forward…

- Clearly central banks use DSGE models because they have appealing theoretical properties, appropriate for policy analysis
  - We have restricted our attention to forecasting properties
- Some recently developed DSGE models, which allow for time variation in the parameters, appear to offer the scope for better predictive densities
- In their absence, ensemble forecast combination offers a reliable methodology for producing well-calibrated forecast densities

niesr