Combining Predictive Densities using Nonlinear Filtering with Applications to US Economics Data

Monica Billio
University of Venice

Roberto Casarin
University of Venice

Francesco Ravazzolo
Norges Bank and BI

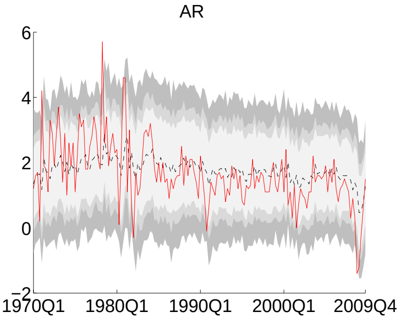
Herman K. van Dijk
Erasmus University Rotterdam

June 2, 2012

Motivation: Density forecasts

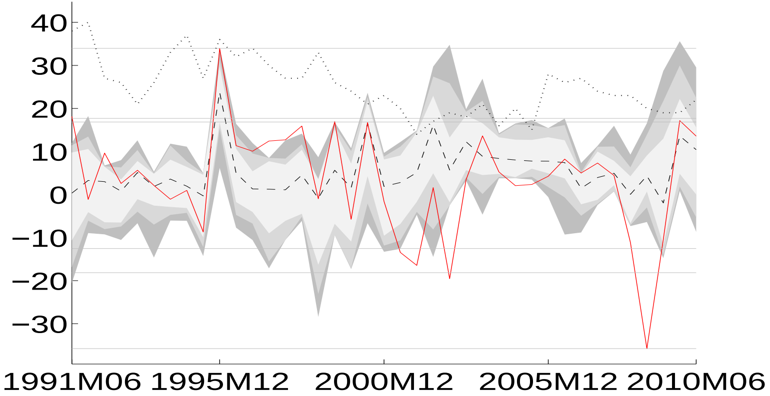
- Complete probability distributions over outcomes provide information helpful for making economic decisions.
- Asset allocation decisions involve higher moments than just first moment.
- Many central banks publish fancharts for forecasts of their variables of interest.

Motivation: US Real GDP Quarterly Growth Rate



Models: 1-quarter ahead forecasts from AR(1) and MS(2)-AR(1).
Simple time series models give large uncertainty in forecasts.

Motivation: Survey Data of US Stock Market (S&P500) Returns



Livingstone survey forecasts for 6-month ahead S&P500 index returns.

Upturn in 1995 well forecasted; downturns around 2001 and in 2009 missed.

Motivation: combination issues

- Averaging as tool to improve forecast accuracy (Barnes (1963), Bates and Granger (1969)).
- Parameter and model uncertainties play an important role (BMA, Roberts (1965)).
- Model performance varies over time, but with some persistence (Diebold and Pauly (1987), Guidolin and Timmermann (2009), Hoogerheide et al. (2010), Gneiting and Raftery (2007)).
- Model set is possible incomplete (Geweke (2009), Geweke and Amisano (2010), Waggoner and Zha (2010)).
- Correlations between forecasts, therefore correlation between weights (Garratt, Mitchell and Vahey (2011)).
- Model performances might differ over quantiles (mixture of predictives).
- Models might perform differently for multiple variables of interest (specific weight for each series, univariate models).

Our contributions: non-Gaussian densities and time varying non-linear weights

- We propose a distributional state-space representation of the predictive densities and of the combination scheme. This representation is general enough to include:
  - Linear and Gaussian models (Granger and Ramanathan (1994)).
  - T-student models (Feng, Villani and Kohn (2009)).
  - Dynamic mixtures of predictives (Huerta, Jiang and Tanner (2003), Villagran and Huerta (2006)).
  - Markov-switching models, copulas, as special cases.

Our contributions: non-Gaussian densities and time varying non-linear weights

- We consider time-varying (and logistic-transformed) weights via convex combinations of the predictive densities (the time-varying weights associated to the different forecasts densities belong to the standard simplex) (Jacobs, Jordan, Nowlan and Hinton (1991)).

- Learning is a possible extension (Diebold and Pauly, (1987)).

- Our weights extend (optimal) least square weights in Granger and Ramanathan (1984), Liang, Zou, Wan and Zhang (2011) and Hansen (2006, 2007).

Applications and results

- We apply our methodology to combine stock index (S&P500) model and survey based density forecasts. Economic and statistical gains. Weight distributions vary over time with with survey based forecasts getting a larger weight in the second of the sample (but some opposite evidence in the tails).

- Model combinations improve the economics gains in our set up.

- Application to GDP growth rate shows the contribution of the learning mechanism in the weights.

- Application to GDP and Inflation still gives large uncertainty in the weights (cannot rule out equal weights).

Previous Papers: Model combinations

- Barnes (1963): the first mention of model combination.

- Roberts (1965): obtained a distribution which includes the predictions from two experts (or models). This distribution is essentially a weighted average of the posterior distributions of two models. This is similar to a Bayesian Model Averaging (BMA) procedure.

- Bates and Granger (1969): seminal paper about combining predictions from different forecasting models.

- Genest and Zidek (1986): pooling of density forecasts.

- Useful reviews: Hoeting et al. (1999) (on BMA with historical perspective), Granger (2006) and Timmermann (2006) (forecasts combination).

Previous Papers: Combination via State-space models

- Granger and Ramanathan (1984): combine the forecasts with unrestricted regression coefficients as weights.
- Diebold and Pauly, (1987) discuss time-varying weights as random walk or with learning.
- Terui and Van Dijk (2002): generalize the least squares model weights by representing the dynamic forecast combination as a state space. In their work the weights are assumed to follow a random walk process.
- Guidolin and Timmermann (2009): introduced Markov-switching weights.
- Hoogerheide et al. (2010) and Groen et al. (2009): robust time-varying weights and accounting for both model and parameter uncertainty in model averaging.
- Hansen (2006, 2007): least squares model averaging and Mallow criteria for optimal restricted [0,1] weights.
- Liang, Zou, Wan and Zhang (2011): theoretical foundation of Bates and Granger.

Notation

- $\mathbf{y}_t \in \mathcal{Y} \subset \mathbb{R}^L$: vector of observable variables;
- $\mathbf{y}_t \sim p(\mathbf{y}_t|\mathbf{y}_{1:t-1})$: conditional forecast density;

- $\tilde{\mathbf{y}}_{k,t} \in \mathcal{Y} \subset \mathbb{R}^L$, with $k = 1, \ldots, K$: a set of one-step-ahead predictors for $\mathbf{y}_t$. (The combination methodology can be extended to multi-step-ahead predictors).
- $\tilde{\mathbf{y}}_{k,t} \sim p(\tilde{\mathbf{y}}_{k,t}|\mathbf{y}_{1:t-1})$, $k = 1, \ldots, K$: conditional density of observable predictive densities.

- $\tilde{\mathbf{y}}_t = \text{vec}(\tilde{Y}_t')$, where $\tilde{Y}_t = (\tilde{\mathbf{y}}_{1,t}, \ldots, \tilde{\mathbf{y}}_{K,t})$.

Previous Combination Methods

Linear pooling

$$p(\mathbf{y}_t|\mathbf{y}_{1:t-1}) = \sum_{k=1}^{K} w_{k,t} p(\tilde{\mathbf{y}}_{k,1:t}|\mathbf{y}_{1:t-1})$$

where $w_{k,t}$ is scalar and it is computed minimizing a loss function.
Mixture of predictives

$$p(\mathbf{y}_t|\mathbf{y}_{1:t-1}) = \sum_{k=1}^{K} g_{k,t}(w_{k,t}|\mathbf{y}_{1:t-1}, \tilde{\mathbf{y}}_{1:t-1}) p(\tilde{\mathbf{y}}_{k,1:t}|\mathbf{y}_{1:t-1})$$

where $g_{k,t}(w_{k,t}|\mathbf{y}_{1:t-1}, \tilde{\mathbf{y}}_{1:t-1})$ is a density.

Combination of Densities (a general representation)

Combination scheme: a probabilistic relation between the density of the observable variable and the predictive densities:

$$p(\mathbf{y}_t|\mathbf{y}_{1:t-1}) = \int_{\tilde{\mathcal{Y}}^{Kt}} p(\mathbf{y}_t|\tilde{\mathbf{y}}_{1:t}, \mathbf{y}_{1:t-1}) p(\tilde{\mathbf{y}}_{1:t}|\mathbf{y}_{1:t-1}) d\tilde{\mathbf{y}}_{1:t}$$

(Conditional dependence structure between $\mathbf{y}_t$ and $\tilde{\mathbf{y}}_{1:t}$: not defined yet).

Combination of Densities (the latent space for the weights)

- $\mathbf{1}_n = (1, \ldots, 1)' \in \mathbb{R}^n$, $\mathbf{0}_n = (0, \ldots, 0)' \in \mathbb{R}^n$

- $\Delta_{[0,1]^n} \subset \mathbb{R}^n$: the set of $\mathbf{w} \in \mathbb{R}^n$ s.t. $\mathbf{w}'\mathbf{1}_n = 1$ and $w_k \geq 0$, $k = 1, \ldots, n$. $\Delta_{[0,1]^n}$ is called the standard n-dimensional simplex and is the latent space.

- $W_t \in \mathcal{W} \subset \mathbb{R}^L \times \mathbb{R}^{KL}$: time-varying weights of the combination scheme. Denote with $w^l_{k,t}$ the k-column and l-row elements of W_t, $\mathbf{w}^l_t = (w^l_{1,t}, \ldots, w^l_{KL,t})'$ s.t. $\mathbf{w}^l_t \in \Delta_{[0,1]^K}$

Latent space: the time series of $[0, 1]$ weights

Weights: interpreted as a discrete p.d.f. over the set of predictors.

Combination of Densities (weight dynamics)

Let $W_t \sim p(W_t|W_{t-1}, \tilde{\mathbf{y}}_{t-\tau:t-1})$ be the density of the time-varying weights, then $p(\mathbf{y}_t|\mathbf{y}_{1:t-1})$ can be written as

$$\int_{\mathcal{Y}^{Kt}} \left(\int_{\mathcal{W}} p(\mathbf{y}_t|W_t, \tilde{\mathbf{y}}_t) p(W_t|\mathbf{y}_{1:t-1}, \tilde{\mathbf{y}}_{1:t-1}) dW_t \right) p(\tilde{\mathbf{y}}_{1:t}|\mathbf{y}_{1:t-1}) d\tilde{\mathbf{y}}_{1:t}$$

where

$$p(W_t|\mathbf{y}_{1:t-1}, \tilde{\mathbf{y}}_{1:t-1}) =$$
$$\int_{\mathcal{W}} p(W_t|W_{t-1}, \tilde{\mathbf{y}}_{t-\tau:t-1}) p(W_{t-1}|\mathbf{y}_{1:t-2}, \tilde{\mathbf{y}}_{1:t-2}) dW_{t-1}$$

Combination of Densities

- Incomplete set of models in $p(\mathbf{y}_t|W_t, \tilde{\mathbf{y}}_t)$ (introducing an error term).
- Multivariate averaging (if $\mathbf{y}_t$ is multivariate).
- Random weights and learning in $p(W_t|\mathbf{y}_{1:t-1}, \tilde{\mathbf{y}}_{1:t-1})$.
- Weights dynamics can account for correlations between forecasts.

Combination of Densities (Example)

Gaussian combination, Logistic-Gaussian Weights with Learning and correlations

$$p(\mathbf{y}_t|W_t,\tilde{\mathbf{y}}_t) \propto \exp\left\{-\frac{1}{2}\left(\mathbf{y}_t - W_t\tilde{\mathbf{y}}_t\right)'\Sigma^{-1}\left(\mathbf{y}_t - W_t\tilde{\mathbf{y}}_t\right)\right\}$$

where the weights are logistic transforms with k elements

$$w_{k,t}^{l} = \frac{\exp\{x_k^l\}}{\sum_{j=1}^{KL}\exp\{x_j^l\}}, \quad \text{with } k = 1,\ldots,KL$$

with $l = 1,\ldots,L$ of the latent process $\mathbf{x}_t$, which has transition

Combination of Densities (Example)

$$p(\mathbf{x}_t|\mathbf{x}_{t-1},\tilde{\mathbf{y}}_{1:t-1}) \propto \exp\left\{-\frac{1}{2}\left(\Delta\mathbf{x}_t - \Delta\mathbf{e}_t\right)' \Lambda^{-1}\left(\Delta\mathbf{x}_t - \Delta\mathbf{e}_t\right)\right\}$$

where $\mathbf{e}_t = \text{vec}(E_t)$, with the elements of $\mathbf{e}_t$ defined by

$$e_{k,t}^{l,d} = (1-\lambda)\sum_{i=1}^{\tau}\lambda^{i-1}(y_{t-i}^{l} - \widehat{y}_{k,t-i}^{l,d})^2$$

- We do not choose between learning and time-varying weights

(Diebold and Pauly (1987), Timmermann (2006)), but combine the two approaches. Λ estimates correlation between weights (extending Clements and Harvey (2011)).

Combination of Densities (Our choice non-linear filtering)

The conditional density $p(\mathbf{y}_t|\mathbf{y}_{t-1})$ can be approximated as follows.

- First, draw j independent values $\mathbf{y}^j_{1:t+1}$, with $j = 1, \ldots, M$ from $p(\tilde{\mathbf{y}}_{s+1}|\mathbf{y}_{1:s})$, with $s = 1, \ldots, t$.

- Conditionally on $\tilde{\mathbf{y}}^j_{1:t+1}$ obtain the particle sets $\Xi^{i,j}_{1:t+1} = \{\mathbf{z}^{i,j}_{1:t+1}, \omega^{i,j}_t\}^N_{i=1}$, with $j = 1, \ldots, M$.

- Simulate $\mathbf{y}^{i,j}_{t+1}$ from $p(\mathbf{y}_{t+1}|\mathbf{z}^{i,j}_{t+1}, \tilde{\mathbf{y}}^j_{t+1})$ and obtain

$$p_{N,M}(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}) = \frac{1}{M}\sum_{j=1}^{M}\sum_{i=1}^{N}\omega^{i,j}_t \delta_{\mathbf{y}^{i,j}_{t+1}}(\mathbf{y}_{t+1})$$

Empirical Applications: GDP and Inflation

- Variables: GDP and inflation measured as PCE deflator.
- Source: Bureau of Economic Analysis.

- Sample: 1960Q1 - 2009Q4.
- Forecasting: 1-step ahead 1980Q1 - 2009Q4.

- Point and density forecasting.
- Individual models: AR and VAR, (2-state) MS AR and VAR.

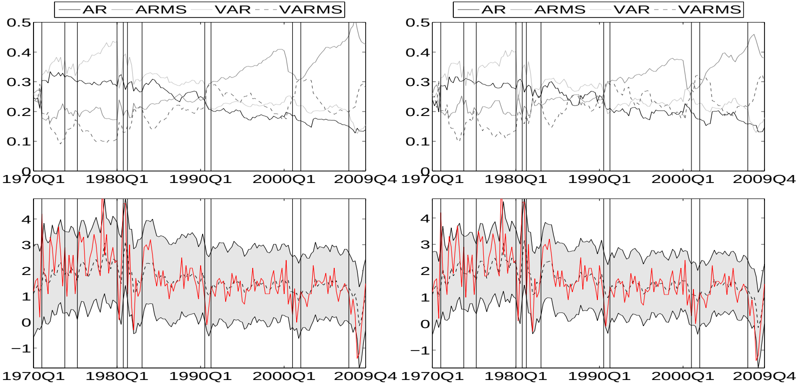
- BMA: based on predictive likelihood (KLIC).
- TVW: time variation.
- TVW(λ, τ): learning with ($\lambda = 0.95$, $\tau = 9$)

Univariate Results (GDP)

	AR	VAR	ARMS	VARMS	BMA	TVW	TVW(λ, τ)
RMSPE	0.882	0.875	0.907	1.000	0.885	0.799	0.691
CW		1.625	1.274	1.587	-0.103	7.185	7.984
LS	-1.323	-1.381	-1.403	-1.361	-2.791	-1.146	-1.151
p_LS		0.337	0.003	0.008	0.001	0.016	0.020
PITS	0.042	0.098	0.164	0.000	0.316	0.468	0.851

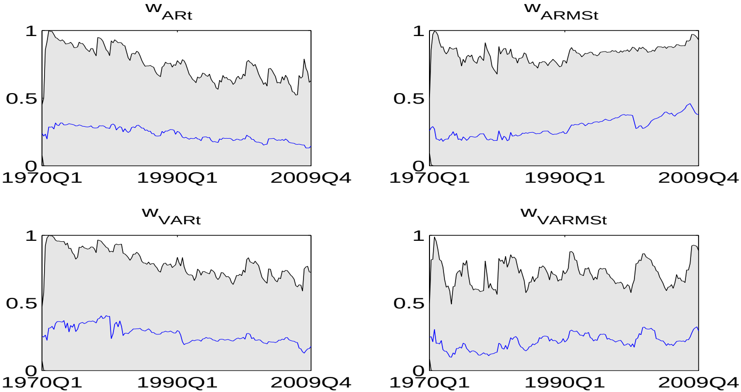
Table: *TVW*: time-varying weights without learning. TVW(λ, τ): time-varying weights with learning mechanism (smoothness parameter $\lambda = 0.95$ and window size $\tau = 9$.)

Weight dynamics: learning effect



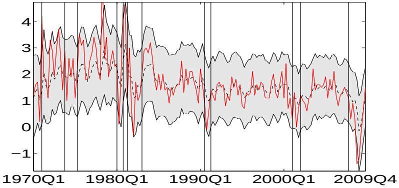
Median weights change over time; learning effect is evident mainly on the tails.

Time-varying weights with learning

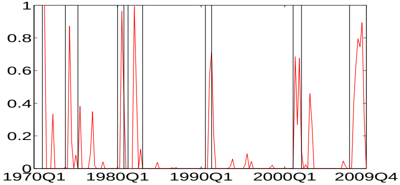


Large uncertainty and equal weights is possible.

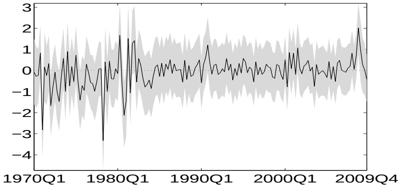
Incompleteness



Fan chart



Turning point predictions



Still large time-variation.

Multivariate Results

	AR	VAR	ARMS	VARMS	BMA	TVW(λ, τ)
GDP						
RMSPE	0.882	0.875	0.907	1.000	0.885	0.718
CW		1.625	1.274	1.587	-0.103	8.554
LS	-1.323	-1.381	-1.403	-1.361	-2.791	-1.012
(p-value)		0.337	0.003	0.008	0.001	0.015
PITS	0.042	0.098	0.164	0.000	0.316	0.958
PCE						
RMSPE	0.385	0.384	0.384	0.612	0.382	0.307
CW		1.036	1.902	1.476	1.234	6.715
LS	-1.538	-1.267	-1.373	-1.090	-1.759	-0.538
(p-value)		0.008	0.024	0.007	0.020	0.024
PITS	0.001	0.000	0.000	0.000	0.000	0.095

Table: Upper table: GDP. Bottom table: PCE.

Empirical Application: Stock Index

- Variables: 6-month Standard & Poor 500 index returns.
- Individual densities: White Noise (WN) and Survey (SR) (nonparametric combination of point forecasts. Parametric: ensemble methodology; Sloughter, Gneiting and Raftery (2010)).
- Source: Livingston Survey Database.

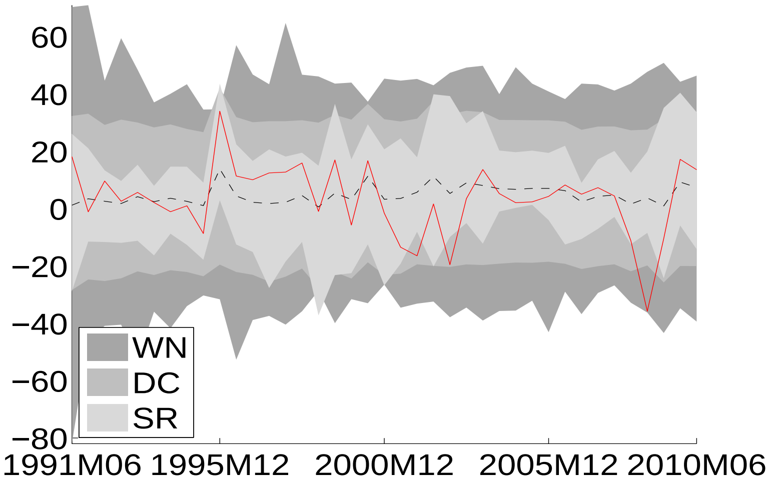
- Sample: 1991M06-2009M12.
- Forecasting: 6-month ahead.

- Point and density forecasting.

- Time-varying weight combinations with learning ($\lambda = 0.95$, $\tau = 9$)

- Risky-risk free power utility investor (no short selling): annualized mean portfolio return, annualized standard deviation, annualized Sharpe ratio and equivalent final values.

Density Combination



Accuracy evaluation 1

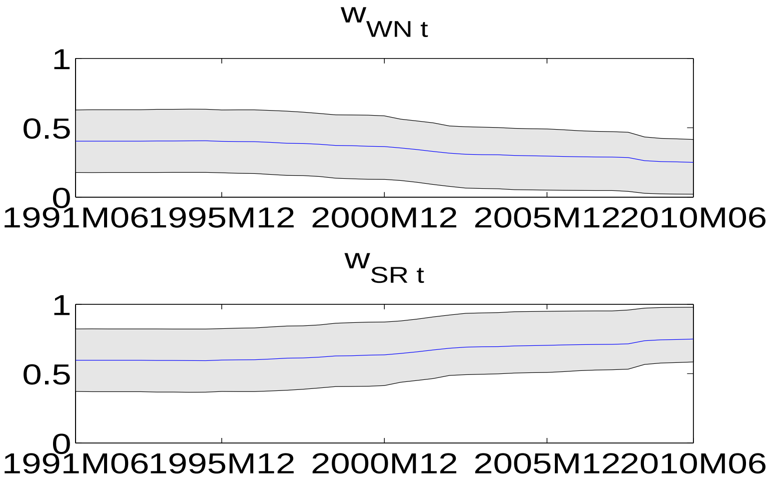
	WN	SR	DC
	Panel A: Statistical accuracy		
RMSPE	12.62	11.23	11.54
SIGN	0.692	0.718	0.692
LS	-3.976	-20.44	-3.880

Accuracy evaluation 2

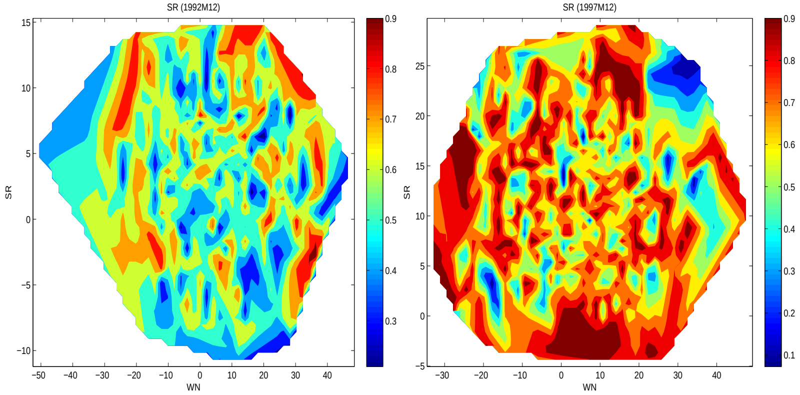
Panel B: Economic analysis									
	$\gamma = 4$			$\gamma = 6$			$\gamma = 8$		
	WN	SR	DC	WN	SR	DC	WN	SR	DC
Mean	5.500	7.492	7.228	4.986	7.698	6.964	4.712	7.603	6.204
St dev	14.50	15.93	14.41	10.62	15.62	10.91	8.059	15.40	8.254
SPR	0.111	0.226	0.232	0.103	0.244	0.282	0.102	0.241	0.280
Utility	-12.53	-12.37	-12.19	-7.322	-7.770	-6.965	-5.045	-6.438	-4.787
r_s	73.1	157.4	254.2	471.5	234.1	671.6	950.9	254.6	1101
r_m	-202.1	-117.8	-20.94	-114.3	-351.7	85.84	3.312	-693.0	153.5
r_b	-138.2	-53.9	43.03	-131.3	-368.8	68.79	-98.86	-795.1	51.32

Results robust to transaction costs.

Weight Dynamics

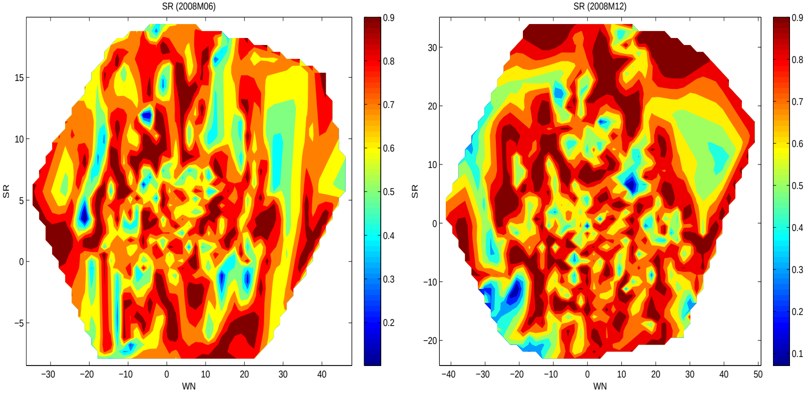


SR weight contours



Model weights differ over quantiles and time.

SR weight contours



Conclusions

- New combination approaches of predictive densities:

  1. Distributional state-space representation and nonlinear Bayesian filtering (Regularised Particle Filter) for the optimal weights estimation.
  2. Nonparametric forecast performance measures for optimal weights estimation.

- Applications to macroeconomics (GDP and PCE) and finance (stock prices).

- Nonlinear combinations with learning outperform (economically and statistically) individual models and BMA.

Future research

- Combining models for turning point forecasts.
- Combining larger set of models, e.g., FAVAR, DSGE.
- Efficient simulation techniques for combining forecast densities defined on high dimensional state space.