Time-varying Combinations of Bayesian Dynamic Models and Equity Momentum Strategies

Herman K. van Dijk

EUR and Norges Bank

(joint with Nalan Baştürk, Stefano Grassi, Lennart Hoogerheide)

Workshop on Forecasting, September 8-9, Deutsche Bundesbank

Introduction and Motivation

Overall goal

- Analyzing the performance of **combinations** of **forecasts** of return models and equity momentum **strategies** in an uncertain dynamic environment with changing data features.

Major challenges

- There exist a **Large number of models**, which potentially explain stylized return features. It is difficult to select the 'best' model.
- Predicted returns from a specific model does not directly lead to a **Practical policy tool** for investors. Selection of the 'best' strategy is not straightforward.
- **Computational issues:** Many potential models (and combinations of these models) are non-linear and non-Gaussian, making use of these models requires efficient computational procedures.
- **Dynamic asset allocation** is a challenging field for practical procedures that also show uncertainty measures.

Four Contributions

1. **Combining flexible model structures**: Methodology to combine different models that capture stylized facts of return distributions (FAVAR-SV and components).
2. **Incorporating policy decision in modeling**: A new dynamic asset-allocation method mixing alternative models and alternative portfolio strategies.
3. **Combination method**: An extended time-varying density combination scheme for model and portfolio strategy mixtures.
4. **Computational tool: M-Filter**: A new filter based on the mixture approximation of the likelihood in each period, aiming to improve efficiency and computing time for density combinations.

Four Contributions

1. **Combining flexible model structures**: Methodology to combine different models that capture stylized facts of return distributions (FAVAR-SV and components).
2. **Incorporating policy decision in modeling**: A new dynamic asset-allocation method mixing alternative models and alternative portfolio strategies.
3. **Combination method**: An extended time-varying density combination scheme for model and portfolio strategy mixtures.
4. **Computational tool: M-Filter**: A new filter based on the mixture approximation of the likelihood in each period, aiming to improve efficiency and computing time for density combinations.

Four Contributions

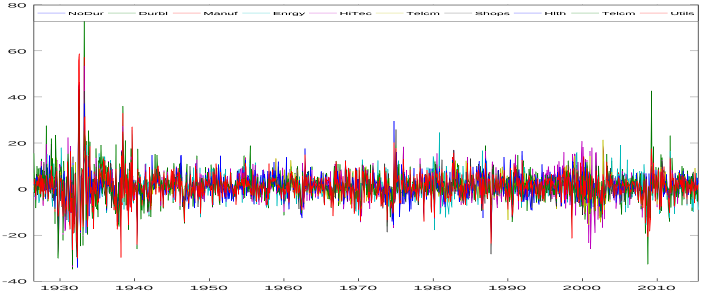
1. **Combining flexible model structures**: Methodology to combine different models that capture stylized facts of return distributions (FAVAR-SV and components).
2. **Incorporating policy decision in modeling**: A new dynamic asset-allocation method mixing alternative models and alternative portfolio strategies.
3. **Combination method**: An extended time-varying density combination scheme for model and portfolio strategy mixtures.
4. **Computational tool: M-Filter**: A new filter based on the mixture approximation of the likelihood in each period, aiming to improve efficiency and computing time for density combinations.

Four Contributions

1. **Combining flexible model structures**: Methodology to combine different models that capture stylized facts of return distributions (FAVAR-SV and components).
2. **Incorporating policy decision in modeling**: A new dynamic asset-allocation method mixing alternative models and alternative portfolio strategies.
3. **Combination method**: An extended time-varying density combination scheme for model and portfolio strategy mixtures.
4. **Computational tool: M-Filter**: A new filter based on the mixture approximation of the likelihood in each period, aiming to improve efficiency and computing time for density combinations.

Monthly percentage returns.

Data: Ten US industry portfolios between 1926M7 and 2015M6.

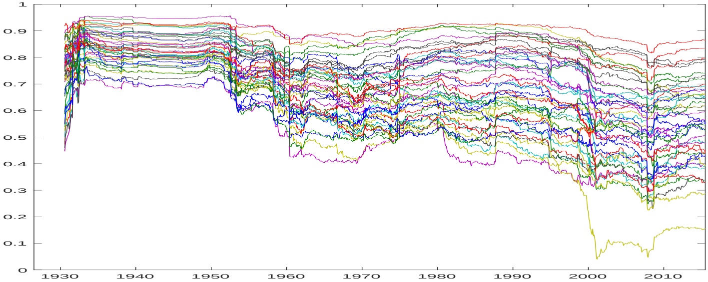


Stylized facts

1. a stationary auto-regressive time-series pattern for all return series.
2. volatility clustering that is common to all series.

Stylized facts - US industry portfolios

Canonical correlations between 45 pairs
Data: Ten US industry portfolios between 1926M7 and 2015M6.

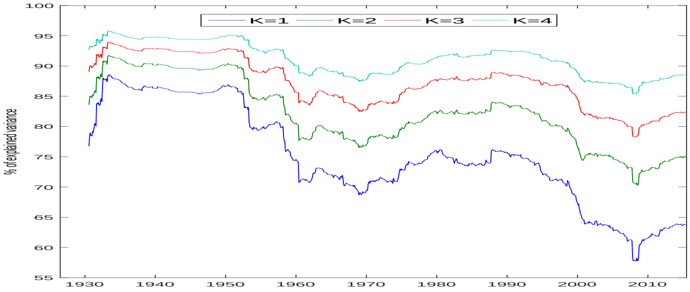


Stylized facts

3 strong cross-section correlation between returns with a time-varying pattern.

Stylized facts - US industry portfolios

Percentage of explained variation by PCA
Data: Ten US industry portfolios between 1926M7 and 2015M6.



Stylized facts

4 total variation in the series can be captured well with one to four components but explained variation (number of common factors) is time-varying.

Short and long-run dynamics in alternative models

General state-space representation for all considered/combined models:

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\beta}\mathbf{x}_t + \Lambda\mathbf{f}_t + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t \sim \mathrm{N}(0, \Sigma_t), \\ \mathbf{f}_t &= \phi_1\mathbf{f}_{t-1} + \ldots + \phi_L\mathbf{f}_{t-L} + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t \sim \mathrm{N}(0, \mathrm{Q}_t), \end{aligned}$$

- **VAR** $\Lambda = 0$, $\mathbf{x}_t$ is the lagged dependent variable, $\boldsymbol{\beta}$ is diagonal.
- **DFM** $\boldsymbol{\beta} = 0$ and a normal distribution for the idiosyncratic and latent disturbances with time-invariant variance-covariance matrices.
- **DFM-SV** DFM with stochastic volatility component in idiosyncratic disturbances, $\boldsymbol{\varepsilon}_t$.
- **FAVAR-SV** FAVAR with stochastic volatility component in idiosyncratic disturbances, $\boldsymbol{\varepsilon}_t$.
- **DFM-SV2** DFM with stochastic volatility component in idiosyncratic and latent disturbances, $\boldsymbol{\varepsilon}_t, \boldsymbol{\eta}_t$.
- **FAVAR-SV2** FAVAR with stochastic volatility component in idiosyncratic and latent disturbances, $\boldsymbol{\varepsilon}_t, \boldsymbol{\eta}_t$.

Contribution: We extend the FAVAR model with one or two SV components, and relate it to relatively simpler models such as the VAR or DFM.

Consider a basic probabilistic combination of densities consisting of

- 1 random variable of interest y, with
- n predictions $\tilde{y} = (\tilde{y}_1, \ldots, \tilde{y}_n)$ from n models.
- Let I denote the information set containing past data and (possibly different) model specifications.

Step 1

Predictive density of random variable y given information set can be calculated from the convolution at the r.h.s.:

$$p(y|I) = \int_{\tilde{Y}} p(y, \tilde{y}|I) d\tilde{y} = \int_{\tilde{Y}} p(y|\tilde{y}, I) p(\tilde{y}|I) d\tilde{y}$$

Step 1 (continued)

$$p(y|I) = \int_{\tilde{Y}} p(y, \tilde{y}|I) d\tilde{y} = \int_{\tilde{Y}} p(y|\tilde{y}, I) p(\tilde{y}|I) d\tilde{y}$$

- Predictive density of a variable of interest y is a weighted average of the conditional density of y given values of predictions $\tilde{y}$, times the marginal density of $\tilde{y}$.
- Formally, $p(y|I)$ is a mixture density where $p(\tilde{y}|I)$ is the mixing density.

Step 2 Specify combination weights $w = (w_1, \ldots, w_n)$ which are unobserved and which connect the predicted values $\tilde{y}$ with the variable to be predicted y as follows:

$$y_t = \widetilde{y}_t' w_t + \epsilon_t \tag{1}$$

with $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$.

$$\begin{aligned} p(y|I) &= \int_{\tilde{Y}} \int_W p(y, \tilde{y}, w|I) d\tilde{y} dw \\ &= \int_{\tilde{Y}} \int_W p(y|w, \tilde{y}, I) p(w|\tilde{y}, I) p(\tilde{y}|I) d\tilde{y} dw \end{aligned}$$

Thus:

- $p(y|w, \tilde{y}, I)$ is a combination density,
- $p(w|\tilde{y}, I)$ is the weight density,
- $p(\tilde{y}|I)$ is the predictive density of all models.

Issue: How to evaluate equations in step 2?

$$p(y|I) = \int_{\tilde{y}} \int_{w} p(y, \tilde{y}, w|I) d\tilde{y} dw \quad (2)$$

$$= \int_{\tilde{y}} \int_{w} p(y|w, \tilde{y}, I) p(w|\tilde{y}, I) p(\tilde{y}|I) d\tilde{y} dw \quad (3)$$

- Suppose the joint density is a normal density. Evaluation is straightforward.
- Suppose that the weight density is markovian dynamic and updating is done in each period with normal densities: Evaluation is straightforward using the Normal/Kalman Filter. However, the weights w are restricted to the unit interval (they are probabilities) and we have a nonlinear transformation using the logistic function to weights that are connected to past predictive performance and to economic weights. Simulation from the weight density is only through indirect sampling methods. Then more involved filtering algorithms labeled Sequential Monte Carlo; see later.

Issue: Comparison BMA and DeCO

- BMA contains true model and for large samples this model is selected
- DeCo allows for model incompleteness. So not only Bayesian learning but also error learning
- BMA has fixed unknown weights.
- DeCo has uncertainty of weights (correlations can be computed) and it has time-varying learning of weights given past predictive performance.

Combining portfolio models and strategies

- Portfolio analysis typically compares realized returns from different strategies and assesses their performance.
- Econometric models yield in accurate predictive densities as input for a portfolio strategy.
- Incorporation of different investment strategies in econometric models is not straightforward.
  This requires a strategy such as mean-variance optimization or a specific utility or loss function

Contribution: Connecting portfolio strategy decisions directly with model comparison and combination, without the need to specify a loss or utility function for the investor.

Considered portfolio strategies

Two equity momentum strategies based on a specific model.

1. *Model Momentum (M.M.):*
   The investor uses the *fitted industry returns* in the past period to go long in assets with the highest posterior mean and to go short in assets with the lowest posterior mean.

   M.M.: Investment decision is based on the model implication directly.

2. *Residual Momentum (R.M.):*
   The investor considers fitted industry returns in the past period for each industry, and invests in the industries with the highest *unexpected returns* during this month, and goes short in stocks with the lowest unexpected returns.

   R.M.: Investment decision is based on surprise/unexpected returns.

Considered portfolio strategies

Two equity momentum strategies based on a specific model.

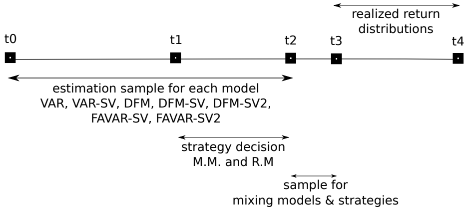
1. *Model Momentum (M.M.):*
   The investor uses the *fitted industry returns* in the past period to go long in assets with the highest posterior mean and to go short in assets with the lowest posterior mean.

   M.M.: Investment decision is based on the model implication directly.

2. *Residual Momentum (R.M.):*
   The investor considers fitted industry returns in the past period for each industry, and invests in the industries with the highest *unexpected returns* during this month, and goes short in stocks with the lowest unexpected returns.

   R.M.: Investment decision is based on surprise/unexpected returns.

Full Bayesian framework to incorporate model and strategy uncertainty



Strategy decision:

- We consider deterministic portfolio strategies $\mathcal{S}_s$ with respect to a single underlying econometric model $\mathcal{M}_m$:

$$\omega_{t,s,m} = g_s(\mathbf{y}_{t-P+1:t}, \varepsilon^{(m)}_{t-P+1:t}),$$

  with past data points $\mathbf{y}_{t-P+1:t}$ and residuals $\varepsilon_{t-P+1:t}$.
  M.M. and R.M. correspond to different deterministic functions $g_s(.)$.
- Given D posterior draws from $\varepsilon^{m,d}_{t-P+1:t}$ for $d = 1, \ldots, D$, we obtain draws from the weight distribution:

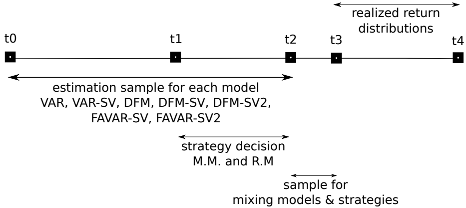
$$\omega^{(d)}_{t,s,m} = g_s(\mathbf{y}_{t-P+1:t}, \varepsilon^{(m,d)}_{t-P+1:t})$$

- In the full Bayesian setting, we also obtain draws from realized returns:

$$r^{\text{real}(d)}_{t+P} = \iota \cdot \mathbf{y}_{t+1:t+P} \cdot \omega^{(d)}_{t,s,m},$$

  where $\mathbf{y}_{t+1:t+P}$ is observed data during the investment period.

Full Bayesian framework to incorporate model and strategy uncertainty



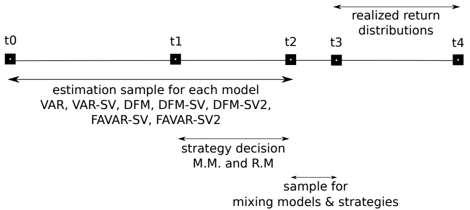
Predicted returns from each strategy:

- Predicted returns from each model and strategy is calculated using the posterior parameter draws.
- We specifically calculate the one period ahead predictive densities for the 'skip period' in portfolio strategies:

$$\tilde{r}^{(d)}_{t+1} = \mathbf{y}^{(m,d)}_{t+1} \cdot \omega^{(d)}_{t,s,m},$$

where $\mathbf{y}^{(m,d)}_{t+1}$ is a draw from the 1 step ahead *forecasts* of returns.

Full Bayesian framework to incorporate model and strategy uncertainty



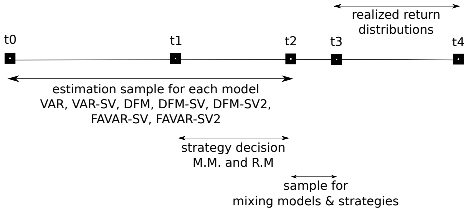
Mixing models and strategies

- We use the one period ahead predictive return distributions to mix models and strategies:

$$f(r_t|I_K) = \sum_{\mathcal{M}_m} \sum_{\mathcal{S}_s} w_{m,s,t} \int_{\mathbb{R}} f(r_t|\tilde{r}_{m,s,t}, I_k) f(\tilde{r}_{m,s,t}|I_k) d\tilde{r}_{m,s,t},$$

- The one-period ahead predictive density corresponds to the ‘skip period’ in standard portfolio construction.

Full Bayesian framework to incorporate model and strategy uncertainty



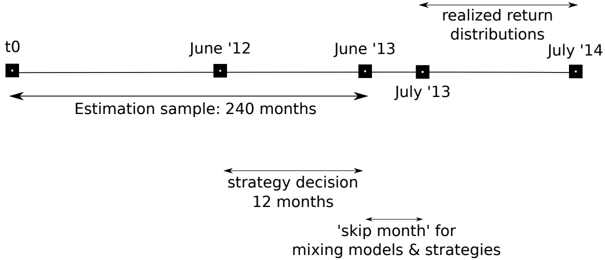
Mixing models and strategies

$$f(r_t|I_K) = \sum_{\mathcal{M}_m} \sum_{\mathcal{S}_s} w_{m,s,t} \int_{\mathbb{R}} f(r_t|\tilde{r}_{m,s,t}, I_k) f(\tilde{r}_{m,s,t}|I_k) d\tilde{r}_{m,s,t},$$

- Difference from a standard model combination scheme: The objective of the combination scheme is to maximize realized return r_t, not the returns of individual stocks.
- An 'optimal' r_t needs to be defined in order to assess the predictive power of each model and strategy combination, hence to infer time-varying weights of these combinations.
- We define this 'optimal return' as the maximum possible return given the information during the skip month t, under the constraint that portfolio weights sum up to 0.

Empirical application

- Data: Ten monthly US industry portfolios between 1926M7 and 2015M6.
- 43 models: VAR, SV, VAR-SV, DFM, DFM-SV, DFM-SV2, FAVAR-SV, FAVAR-SV2 with different autocorrelation structures (number of factors, number of AR lags in the latent variable).
- 2 investment strategies (M.M. and R.M.) for each model.
- In total, we have 86 components that can potentially be compared/combined.
- Investment decisions are made once a year, strategies are based on the return performance during the last 12 months.



Realized returns from different models

	(K, L)	Model Momentum Mean	Vol.	S.R.	L.L.	Residual Momentum Mean	Vol.	S.R.	L.L.
VAR-N	–	0.02	**5.0**	0.005	**-24.1**	**0.09**	5.8	0.015	-35.0
SV	–	**0.10**	**5.1**	0.019	-34.7	**0.11**	**5.6**	0.019	**-26.0**
VAR-SV	–	**0.12**	**4.5**	**0.028**	-20.2	**0.13**	5.8	**0.021**	-37.4
DFM-N	(1,1)	-0.04	**4.9**	-0.009	**-20.0**	**0.13**	**5.7**	**0.023**	-34.4
	(1,2)	-0.04	**4.9**	-0.009	**-20.0**	**0.13**	**5.7**	**0.022**	-34.4
	(2,1)	-0.13	**5.2**	-0.024	**-25.4**	**0.10**	**5.6**	0.017	-34.0
	(2,2)	-0.11	**5.2**	-0.020	**-24.2**	**0.10**	**5.6**	0.017	-34.1
	(3,1)	-0.14	**5.4**	-0.027	**-23.7**	**0.09**	**5.5**	0.017	-33.7
	(3,2)	-0.08	**5.4**	-0.016	**-23.3**	0.08	**5.4**	0.015	-33.1
	(4,1)	-0.07	**5.5**	-0.013	-26.7	**0.10**	**5.4**	0.018	-31.3
	(4,2)	-0.05	**5.5**	-0.009	-27.4	**0.12**	**5.4**	**0.022**	-31.1
DFM-SV	(1,1)	0.04	**5.0**	**0.007**	**-20.0**	**0.11**	5.8	0.019	-37.1
	(1,2)	0.04	**5.0**	**0.008**	**-20.0**	**0.10**	5.8	0.018	-37.1
	(2,1)	-0.04	**5.2**	-0.009	**-22.0**	**0.15**	**5.7**	**0.026**	-36.3
	(2,2)	-0.05	**5.2**	-0.009	**-22.0**	**0.15**	**5.7**	**0.027**	-36.6
	(3,1)	0.00	**5.2**	0.000	**-21.2**	**0.14**	**5.4**	**0.026**	-33.0
	(3,2)	0.03	**5.2**	0.005	**-20.8**	**0.16**	**5.4**	**0.030**	-32.8
	(4,1)	**0.12**	**5.4**	**0.023**	**-20.8**	0.05	**5.4**	0.009	-31.8
	(4,2)	**0.12**	**5.4**	**0.023**	**-21.7**	0.06	**5.4**	0.011	-31.1
DFM-SV2	(1,1)	0.07	**4.6**	0.014	**-18.2**	0.06	**5.5**	0.010	-37.4
	(1,2)	0.07	**4.6**	0.014	**-18.2**	0.06	**5.5**	0.010	-37.4
	(2,1)	-0.01	**4.8**	-0.002	**-22.8**	0.08	**5.5**	0.015	-37.4
	(2,2)	-0.02	**4.8**	-0.003	**-22.8**	**0.09**	**5.5**	0.016	-37.4
	(3,1)	0.02	**5.0**	0.005	-27.1	-0.02	**5.5**	-0.003	-37.4
	(3,2)	0.03	**5.0**	0.006	-27.1	-0.02	**5.5**	-0.003	-37.4
	(4,1)	0.07	**5.7**	0.013	-32.3	0.00	**5.2**	0.000	-37.4
	(4,2)	0.07	**5.7**	0.013	-32.3	0.00	**5.2**	0.000	-37.4

Bold values indicate better performance compared to standard momentum.

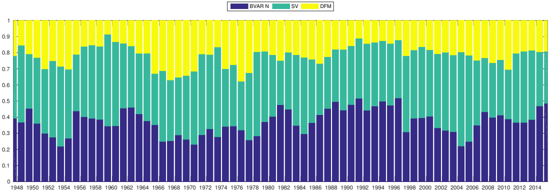
Realized returns from different models

	(K, L)	Model Momentum Mean	Vol.	S.R.	L.L.	Residual Momentum Mean	Vol.	S.R.	L.L.
FAVAR-SV	(1,1)	0.08	**4.6**	0.018	**-18.3**	0.06	**5.5**	0.011	-37.4
	(1,2)	0.08	**4.6**	0.018	**-18.3**	0.06	**5.5**	0.011	-37.4
	(2,1)	-0.03	**4.9**	-0.005	**-23.1**	0.08	**5.5**	0.015	-37.4
	(2,2)	-0.03	**4.9**	-0.006	**-23.5**	**0.09**	**5.5**	0.016	-37.4
	(3,1)	**0.09**	**5.0**	0.018	**-25.3**	-0.02	**5.5**	-0.005	-37.4
	(3,2)	0.08	**5.0**	0.017	**-25.7**	-0.02	**5.5**	-0.004	-37.4
	(4,1)	0.08	**5.7**	0.014	-32.3	0.03	**5.2**	0.005	-37.4
	(4,2)	0.08	**5.7**	0.015	-32.3	0.02	**5.2**	0.005	-37.4
FAVAR-SV2	(1,1)	**0.09**	**4.6**	0.019	**-18.3**	0.06	**5.5**	0.011	-37.4
	(1,2)	0.08	**4.6**	0.018	**-18.3**	0.06	**5.5**	0.011	-37.4
	(2,1)	-0.03	**4.9**	-0.005	**-23.5**	**0.09**	**5.5**	0.016	-37.4
	(2,2)	-0.03	**4.9**	-0.005	**-23.8**	0.08	**5.5**	0.015	-37.4
	(3,1)	0.08	**5.0**	0.017	**-25.6**	-0.03	**5.5**	-0.005	-37.4
	(3,2)	0.08	**5.0**	0.017	**-25.3**	-0.02	**5.5**	-0.004	-37.4
	(4,1)	0.08	**5.7**	0.014	-32.3	0.03	**5.2**	0.005	-37.4
	(4,2)	0.08	**5.7**	0.014	-32.3	0.03	**5.2**	0.005	-37.4

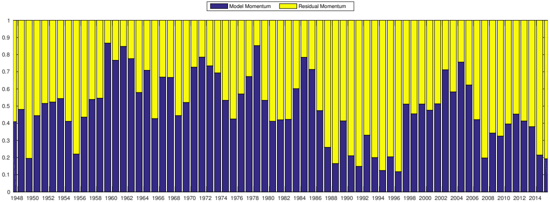
Bold values indicate better performance compared to standard momentum.

- Residual momentum typically leads to higher returns.
- Model momentum typically leads to lower risk.
- Model choice, together with the autocorrelation patterns (K,L) are important.
- SV component in idiosyncratic errors is important, particularly for M.M..
- SV component in latent errors does not contribute to realized returns substantially.

Three basic models: VAR-N (AR pattern), SV (time-varying volatility), DFM (cross-sectional correlation).
Two investment strategies: M.M. and R.M.



Cumulative model weights (post. mean)



Cumulative strategy weights (post. mean)

- Model and strategy weights vary over time.
- Time variation is particularly relevant for strategies.

Mixture of three basic models and two investment strategies

Three basic models: VAR-N (AR pattern), SV (time-varying volatility), DFM (cross-sectional correlation).
Two investment strategies: M.M. and R.M.

Model	Strategy	Mean	Vol.	S.R.	L.L.
Mixture of basic models and two strategies					
VAR-N & SV & DFM-N(4,2)	M.M. & R.M.	0.10 (0.01,0.18)	3.9 (3.6,4.2)	0.025 (0.002,0.047)	-23.0 (-28.8,-17.5)
Mixture of strategies per model					
VAR-N	M.M. & R.M.	0.09 (-0.03,0.20)	4.7 (4.0.4,5)	0.019 (-0.007,0.043)	-32.6 (-35.6,-20.9)
SV	M.M. & R.M.	0.13 (-0.02,0.28)	4.3 (3.9,4.6)	0.032 (-0.005,0.065)	-22.2 (-29.9,-16.1)
DFM-N(4,2)	M.M. & R.M.	0.03 (-0.12,0.17)	4.3 (4.0,4.7)	0.006 (-0.028,0.041)	-24.4 (-31.1,-16.8)

Standard momentum: mean 0.09, volatility 5.7, Sharpe ratio 0.02 and largest loss -26.2.

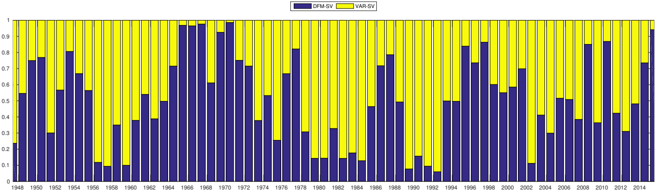
- Substantial reduction in variances both due to the mixture of strategies and due to mixture of models.
- Mean returns have large uncertainty, hence the mixture of models and strategies also have high uncertainty in mean returns.

Mixture of two flexible models and a mixture of investment strategies

Two flexible models:

- VAR-SV.
- DFM-SV with 1-4 factors and 1-2 AR lags.

Two investment strategies: M.M. and R.M.



Mixture of two flexible models and a mixture of investment strategies

Two flexible models:

- VAR-SV.
- DFM-SV with 1-4 factors and 1-2 AR lags.

Two investment strategies: M.M. and R.M.

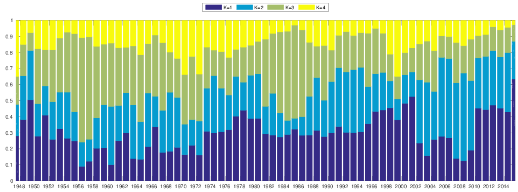
Model	Strategy	Mean	Vol.	S.R.	L.L.
Mixture of two flexible models and strategies					
VAR-SV & DFM-SV(1:4,1:2)	M.M. & R.M.	0.15 (0.08, 0.22)	3.7 (3.5, 3.9)	0.041 (0.021, 0.061)	-21.6 (-26.4, -16.4)
Mixture of strategies per model					
VAR-SV	M.M. & R.M.	0.23 (0.11, 0.35)	4.5 (4.2, 4.9)	0.051 (0.024, 0.080)	-37.2 (-37.3, -36.8)
DFM-SV(1:4,1:2)	M.M. & R.M.	0.06 (0.00, 0.12)	3.4 (3.2, 3.5)	0.018 (0.000, 0.036)	-14.4 (-20.1, -11.0)

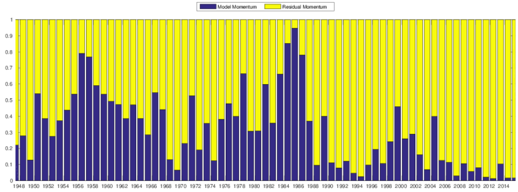
Standard momentum: mean 0.09, volatility 5.7, Sharpe ratio 0.02 and largest loss -26.2.

- Mixture of models and strategies again improve risk measures in general.
- Mean returns are in general ‘higher’ than those of the mixture of standard models.
- The combination of models and strategies are useful, but the choice of underlying models should also be taken into account.

Mixture of very flexible parametric model and investment strategies

- Flexible model: FAVAR-SV with 1-4 factors, 1-2 lags in the factor equation.
- Two investment strategies: M.M. and R.M.





Mixture of very flexible parametric model and investment strategies

- Flexible model: FAVAR-SV with 1-4 factors, 1-2 lags in the factor equation.
- Two investment strategies: M.M. and R.M..

Model	Strategy	Mean	Vol.	S.R.	L.L.
Mixture of models and two strategies					
FAVAR-SV(1:4, 1:2)	M.M. & R.M.	0.18 (0.14, 0.22)	4.5 (4.5, 4.6)	0.039 (0.031, 0.048)	-34.8 (-35.0, -34.6)
Mixture of strategies per model					
FAVAR-SV(1, 1)	M.M. & R.M.	0.11 (0.02, 0.19)	4.5 (4.4, 4.6)	0.024 (0.004, 0.042)	-33.8 (-34.0, -33.1)
FAVAR-SV(1, 2)	M.M. & R.M.	0.11 (0.02, 0.19)	4.5 (4.4, 4.6)	0.023 (0.004, 0.042)	-34.2 (-34.4, -33.6)
FAVAR-SV(2, 1)	M.M. & R.M.	0.14 (0.05, 0.22)	5.1 (5.0, 5.2)	0.027 (0.010, 0.043)	-37.1 (-37.2, -36.9)
FAVAR-SV(2, 2)	M.M. & R.M.	0.14 (0.05, 0.22)	5.1 (5.0, 5.2)	0.027 (0.010, 0.044)	-37.1 (-37.2, -36.8)
FAVAR-SV(3, 1)	M.M. & R.M.	0.15 (0.07, 0.25)	4.7 (4.5, 4.9)	0.033 (0.014, 0.054)	-34.1 (-34.3, -34)
FAVAR-SV(3, 2)	M.M. & R.M.	0.14 (0.05, 0.25)	4.7 (4.6, 4.9)	0.031 (0.011, 0.052)	-34.4 (-34.5, -34.2)
FAVAR-SV(4, 1)	M.M. & R.M.	0.11 (0.02, 0.20)	5.1 (5.0, 5.2)	0.022 (0.004, 0.040)	-31.3 (-31.8, -31.1)
FAVAR-SV(4, 2)	M.M. & R.M.	0.12 (0.03, 0.21)	5.1 (5.0, 5.2)	0.023 (0.005, 0.040)	-31.5 (-32.4, -31.3)

Standard momentum: mean 0.09, volatility 5.7, Sharpe ratio 0.02 and largest loss -26.2.

General empirical conclusions

- Flexible model mixtures lead to higher means and Sharpe ratios than mixtures of basic model structures where one component fits very poorly.
  Thus, choice of the model set in the sense of choosing the number of components in a mixture is important for effective momentum strategies.
- A mixture of our two strategies leads, in particular, to better risk features. Here the information of complete densities plays an important role.
- There is no clear optimal result in terms of return and risk features.
  Alternative mixtures of models and strategies in different time periods may be effective in improving returns and risk.
- The time-varying nature of the results remains robust over many different alternatives.

MitISEM Filter (M-Filter)

- The flexible model and strategy combination (density combinations) is estimated using a non-linear and non-Gaussian state space model.
- Such a flexible combination brings robustness and computing time challenges.

> Contribution: M-filter A novel non-linear non-Gaussian filter which uses mixtures of student-t distributions.

Main properties:

- An adjusted particle filter where the proposal density for the non-linear non-Gaussian state variable is based on a Mixture of Student-t densities (MitISEM) with an unknown number of components.
- The proposal for the state variable is constructed at every filtering step and for each time period.

Two advantages:

- Possibility to handle complex posterior distributions using flexible student-t mixtures.
- Updated proposal densities at every time period t avoids particle depletion. The step of resampling in the particle filter is not required.

MitISEM Filter (M-Filter)

A generic state space model and particle filter (PF) recursions:

$$\mathbf{y}_t = m_t(\boldsymbol{\alpha}_t, \boldsymbol{\varepsilon}_t),$$
$$\boldsymbol{\alpha}_t = h_t(\boldsymbol{\alpha}_{t-1}, \boldsymbol{\eta}_t),$$

where $\mathbf{y}_t$ are the observations, $\boldsymbol{\alpha}_t$ are the state variables and $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_t$ are mutually independent errors.

- State variables $\boldsymbol{\alpha}_T = \{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_t\}$ are typically unobserved.
- The object of interest is the joint conditional distribution:

$$p(\boldsymbol{\alpha}_T | \mathbf{y}_T, \hat{\theta}) = \frac{p(\boldsymbol{\alpha}_T, \mathbf{y}_T | \hat{\theta})}{p(\mathbf{y}_T | \hat{\theta})},$$

where the $p(\mathbf{y}_T | \hat{\theta})$ is the likelihood of the state space model.

MitISEM Filter (M-Filter)

Background: Particle filter steps conditional on the estimated parameters:

1) **Initialization**. Draw the initial particles $\boldsymbol{\alpha}_0^{(j)} \sim p(\boldsymbol{\alpha}_0)$ and set the weights $\mathrm{W}_0^{(j)} = 1$ for $j = 1, \ldots, M$.
2) **Recursion**. For $t = 1, \ldots, T$

   a.) **Forecasting $\boldsymbol{\alpha}_t$**. Draw $\tilde{\boldsymbol{\alpha}}_t^{(j)}$ from the density $g_t(\tilde{\boldsymbol{\alpha}}_t^{(j)} | \boldsymbol{\alpha}_{t-1}^{(j)}, \hat{\theta})$ and define the importance weights as:

$$\omega_t^{(j)} = \frac{p(\tilde{\boldsymbol{\alpha}}_t^{(j)} | \boldsymbol{\alpha}_{t-1}^{(j)}, \hat{\theta})}{g_t(\tilde{\boldsymbol{\alpha}}_t^{(j)} | \boldsymbol{\alpha}_{t-1}^{(j)}, \hat{\theta})}.$$

   b.) **Forecasting $\mathbf{y}_t$**. Define the incremental weights: $\tilde{\omega}_t^{(j)} = p(\mathbf{y}_t | \tilde{\boldsymbol{\alpha}}_t^{(j)}, \hat{\theta}) \omega_t^{(j)}$.

3) **Updating**. Define the normalized weights $\tilde{\mathrm{W}}_t^{(j)}$ and an approximation of $\mathrm{E}[h_t(\tilde{\boldsymbol{\alpha}}_t) | \mathbf{y}_{1:t}, \theta]$

$$\tilde{\mathrm{W}}_t^{(j)} = \frac{\tilde{w}_t^{(j)} \mathrm{W}_{t-1}^{(j)}}{\frac{1}{M} \sum_{j=1}^{M} \tilde{w}_t^{(j)} \mathrm{W}_{t-1}^{(j)}}., \qquad \tilde{h}_{t,M} = \frac{1}{M} \sum_{j=1}^{M} h_t(\tilde{\boldsymbol{\alpha}}_t^{(j)}).$$

4) **Selection**. Resample the particle via e.g. multinomial resampling.
5) **Likelihood Approximation**. The approximation of the log likelihood function is given by:

$$\log \hat{p}(\mathbf{y}_{1:T} | \hat{\theta}) = \sum_{t=1}^{T} \log \left(\frac{1}{M} \sum_{j=1}^{M} \tilde{w}_t^{(j)} \mathrm{W}_{t-1}^{(j)} \right).$$

M-filter steps

At each time t we construct the importance density $g_t(\tilde{\boldsymbol{\alpha}}_t^{(j)}|\boldsymbol{\alpha}_{t-1}^{(j)})$ around the target density $p(\mathbf{y}_t|\tilde{\boldsymbol{\alpha}}_t^{(j)})p(\tilde{\boldsymbol{\alpha}}_t^{(j)}|\boldsymbol{\alpha}_{t-1}^{(j)})$.

1) **Initialization**. Draw $\tilde{\boldsymbol{\alpha}}_0^{(j)} \sim p(\boldsymbol{\alpha}_0)$ for $j = 1, \ldots, M$.

2) **Recursion**. For $t = 1, \ldots, T$ construct $\tilde{g}_t(\tilde{\boldsymbol{\alpha}}_t^{(j)}|\boldsymbol{\alpha}_{t-1}^{(j)})$ using the MitISEM procedure:

    a.) **Initialization**: Simulate draws $\tilde{\boldsymbol{\alpha}}_t^{(j)}$ from a 'naive' proposal distribution with density $g_n(\cdot)$ (e.g. a Student-t with v degrees of freedom). Using the target density:

    $$p(\mathbf{y}_t|\tilde{\boldsymbol{\alpha}}_t^{(j)})p(\tilde{\boldsymbol{\alpha}}_t^{(j)}|\boldsymbol{\alpha}_{t-1}^{(j)}),$$

    update the the mode and scale of the proposal density using the IS weighted EM algorithm.

    b.) **Adaptation**: Update parameters of the proposal density using the MitISEM procedure.

3) **Draws**. Draws $\tilde{\boldsymbol{\alpha}}_t^{(j)}$ from the constructed density $\tilde{g}_t(\tilde{\boldsymbol{\alpha}}_t^{(j)}|\boldsymbol{\alpha}_{t-1}^{(j)})$ and approximate $\mathrm{E}[h_t(\boldsymbol{\alpha}_t)|\mathbf{y}_{1:T}]$ by:

$$\boldsymbol{\alpha}_t = \frac{1}{M} \sum_{j=1}^{M} h(\tilde{\boldsymbol{\alpha}}_t^{(j)}).$$

4) **Likelihood Approximation**. The approximation of the log likelihood:

$$\log \hat{p}(\mathbf{y}_{1:T}) = \sum_{t=1}^{T} \log \left(\frac{1}{M} \sum_{j=1}^{M} \tilde{w}_t^{(j)} \right).$$

where $\tilde{w}_t^{(j)}$ are the weights at time t.

Approximation and speed comparisons between M–Filter and other methods

We examine a DFM model with $K = \{2, 4, 6, 10\}$ factors, $T = 100$ and $N = 20$ in $I = 100$ replications.

# Factors:	2			4			2	4
	LB	Bias	Var	LB	Bias	Var	*Time*	
KF	0.00	1.00	1.00	0.00	1.00	1.00	0.011	0.012
PF	-77.42	1.15	1.33	-145.49	1.15	1.32	708.790	811.730
APF	-39.98	1.03	1.05	-164.80	1.05	1.05	836.690	878.128
MF	-23.23	1.01	1.02	-23.39	1.00	1.01	106.330	138.178
# Factors:	6			10			6	10
	LB	Bias	Var	LB	Bias	Var	*Time*	
KF	0.00	1.00	1.00	0.00	1.00	1.00	0.020	0.021
PF	-193.74	1.16	1.31	-333.33	1.27	1.65	861.100	897.860
APF	-309.26	1.07	1.12	-568.18	1.08	1.18	953.720	1011.210
MF	-16.97	1.03	1.03	-112.68	1.02	1.03	213.200	402.820

- Kalman Filter (KF), Bootstrap Particle Filter (PF), Auxiliary Particle Filter (APF) and M–Filter (MF) results with 50000 particles.
- For DFM, KF is the optimal filter, hence the natural benchmark for comparing the filters.
- **LB**: Likelihood Bias relative the Kalman Filter.
- **Bias**: Absolute errors in state estimates $1/I \sum_{i=1}^{I} \mid \tilde{\alpha}_{t,i} - \alpha_{t,i} \mid$ relative to KF.
- **Var**: Variability defined as $1/I \sum_{i=1}^{I} (\tilde{\alpha}_{t,i} - \alpha_{t,i})^2$ relative to KF.
- **Time:** Computational time in seconds.

Approximation and speed comparisons between M-filter and other methods

We examine three cases of structural breaks in AR(1) models. Model set (for simulation and DECO M-Filter) includes five models:

$$\tilde{y}_{1,t} = 0.1 + 0.1\tilde{y}_{1,t-1} + \varepsilon_t \qquad \varepsilon_t \sim \mathrm{N}(0,1) \qquad \tilde{y}_{4,t} = 0.4 + 0.4\tilde{y}_{4,t-1} + \varepsilon_t \qquad \varepsilon_t \sim \mathrm{N}(0,1)$$
$$\tilde{y}_{2,t} = 0.2 + 0.2\tilde{y}_{2,t-1} + \varepsilon_t \qquad \varepsilon_t \sim \mathrm{N}(0,1) \qquad \tilde{y}_{5,t} = 0.5 + 0.5\tilde{y}_{5,t-1} + \varepsilon_t \qquad \varepsilon_t \sim \mathrm{N}(0,1)$$
$$\tilde{y}_{3,t} = 0.3 + 0.3\tilde{y}_{3,t-1} + \varepsilon_t \qquad \varepsilon_t \sim \mathrm{N}(0,1)$$

Case 1: One model has weight 1 for all t:

$$y_{1:t} = \tilde{y}_{1,1:t} + \eta_t \qquad \eta_t \sim \mathrm{N}(0, 0.05)$$

Case 2: One switch at $t = 101$ from $\tilde{y}_1$ to $\tilde{y}_5$

$$y_{1:100} = \tilde{y}_{1,1:100} + \eta_t \qquad \eta_t \sim \mathrm{N}(0, 0.05)$$
$$y_{101:t} = \tilde{y}_{5,101:t} + \eta_t \qquad \eta_t \sim \mathrm{N}(0, 0.05)$$

Case 3: Two switches at $t = 101$ $(\tilde{y}_1 \to \tilde{y}_5)$ and $t = 151$ $(\tilde{y}_5 \to \tilde{y}_3)$.

$$y_{1:100} = \tilde{y}_{1,1:100} + \eta_t \qquad \eta_t \sim \mathrm{N}(0, 0.05)$$
$$y_{101:150} = \tilde{y}_{5,101:150} + \eta_t \qquad \eta_t \sim \mathrm{N}(0, 0.05)$$
$$y_{151:t} = \tilde{y}_{3,151:t} + \eta_t \qquad \eta_t \sim \mathrm{N}(0, 0.05)$$

Approximation and speed comparisons between M-filter and other methods

Comparison of different filters for structural breaks in AR(1) models.

Model	Case 1 (no break)			Case 2 (one break)			Case 3 (two breaks)		
	Bias	Variance	Time	Bias	Variance	Time	Bias	Variance	Time
KF	1.000	1.000	0.007	1.000	1.000	0.007	1.000	1.000	0.007
PF	0.019	0.001	58.483	**0.057**	0.052	58.483	0.123	0.202	58.483
APF	**0.008**	**0.001**	68.015	0.061	0.081	68.015	0.091	0.077	68.015
M-Filter	0.059	0.007	**39.993**	0.065	**0.039**	**40.676**	**0.079**	**0.067**	**41.180**

Bias and variability are reported relative to the KF. The results are obtained from $I = 100$ iterations, with 50,000 particles for PF, APF, and M-Filter.

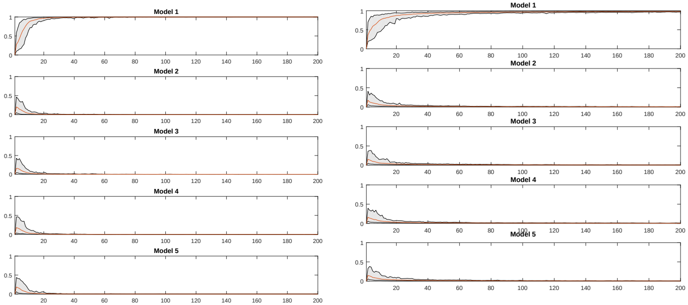
- **Case 1:** One model has weight 1 for all t: Gains from M-filter is minimal.
- **Cases 2:** Bias difference between PF, APF and M-Filter reduces, M-filter has the smallest variance.
- **Cases 3:** M-filter performs best due to the higher number of breaks and the adaptation of the density at each time period.
- In all cases, M-filter is computationally more efficient compared to all methods but the Kalman Filter.
- We next compare the weights obtained by APF and M-Filter visually.

Approximation and speed comparisons between M-filter and other methods

Model weights for Case 1: Model 1 has weight 1 for all t

APF weights M-filter weights

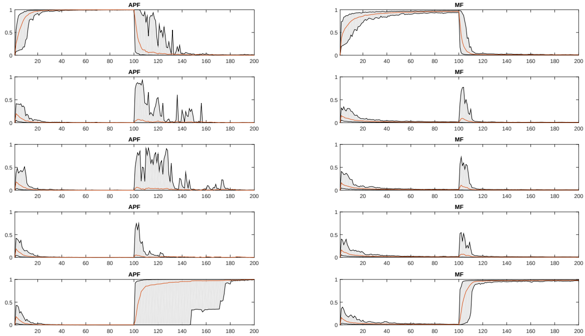
(posterior mean and 95% credible intervals)



- M-Filter relatively slowly adjusts to model 1 weight of 1.

Model weights for Case 2: Switch between model 1 $\rightarrow$ model 5.

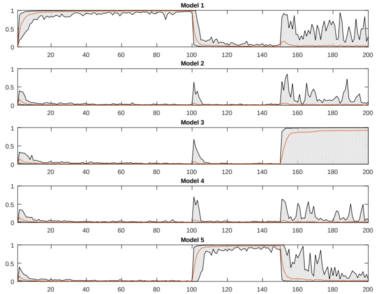
APF weights M-filter weights
(posterior mean and 95% credible intervals)

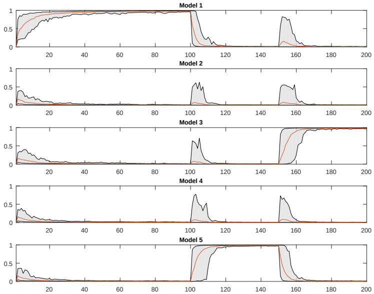


- M-Filter is faster in picking up the ‘break’ due to updated candidate at each time period.

Model weights for Case 2: Switch between model 1 $\rightarrow$ model 5 $\rightarrow$ model 3.

APF weights M-filter weights
(posterior mean and 95% credible intervals)





- M-Filter is faster in picking up the ‘breaks’ (particularly the second break).
- This is due to updated candidate at each time period.

References

- Baştürk N, Grassi S, Hoogerheide L, Van Dijk H K. 2016. Parallelization Experience with Four Canonical Econometric Models Using ParMitISEM. *Econometrics* **4**: 1–11.
- Bernanke S B, Boivin J, Eliasz P. 2005. Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach. *The Quarterly Journal of Economics* **120**: 387–422.
- Blitz D, Huij J, Martens M. 2011. Residual Momentum. *Journal of Empirical Finance* **18**: 506–521.
- Casarin R, Grassi S, Ravazzolo F, Van Dijk H K. 2016. Dynamic Predictive Density Combinations for Large Data Sets in Economics and Finance. Technical Report 2015–084/III, Tinbergen Institute.
- Hoogerheide L, Opschoor A, Van Dijk H K. 2012. A Class of Adaptive Importance Sampling Weighted EM Algorithms for Efficient and Robust Posterior and Predictive Simulation. *Journal of Econometrics* **171**: 101–120.
- Jacobs R A, Jordan M I, Nowlan S J, Hinton G E. 1991. Adaptive mixtures of local experts. *Journal of Neural Computation* **3**: 79–87.
- Stock J H, Watson W M. 2005. Implications of Dynamic Factor Models for VAR Analysis. Technical report, NBER Working Paper No. 11467.