Preliminary Draft. Please do not cite, circulate or quote without the authors' permission

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

N. Baştürk[1], A. Borowska[3,4], S. Grassi[2], L. Hoogerheide[3,4], and H. K. van Dijk[3,5,6]

[1]Maastricht University
[2]University of Rome, Tor Vergata
[3]Tinbergen Institute
[4]VU University Amsterdam
[5]Erasmus University Rotterdam
[6]Norges Bank

October 28, 2017

Abstract

A novel dynamic asset-allocation approach is proposed where portfolios as well as portfolio strategies are updated at every decision period based on their past performance. A general class of models is specified, combining a dynamic factor model and a vector autoregressive model, where we also allow for stochastic volatility. A Bayesian strategy combination is introduced to combine two model-based strategies, a model momentum strategy based on fitted returns and a residual momentum strategy. This extends the mixture of experts analysis by allowing the strategy weights to be interdependent, time-dependent and incomplete. The presented estimation approach, originating from the forecast combination literature, relies on the implied state space structure of the joint model for the time series and strategy weights. Given the complexity of the resulting non-linear and non-Gaussian structure a novel

*This paper should not be reported as representing the views of Norges Bank. The views expressed are those of the authors and do not necessarily reflect those of Norges Bank.

†We would like to thank the participants at the 7th European Seminar on Bayesian Econometrics Annual Workshop in Venice, and the Sequential Monte Carlo 2017 Workshop in Uppsala, Norges Bank Conference on 'Big data, machine learning and the macroeconomy' in Oslo and the 8th European Seminar on Bayesian Econometrics Annual Workshop in Maastricht for their insightful comments.

and efficient particle filter is introduced, based on mixtures of Student's t distributions. Using US industry portfolio returns over almost a century of monthly data, our empirical results indicate that time-varying combinations of flexible models from our developed class with two momentum strategies outperform competing models in terms of mean returns and Sharpe ratios, as well as reduced volatility and the largest loss. The latter result demonstrates the usefulness of the proposed methodology from risk management perspective.

1 Introduction

Traditional factor models rely on macro or firm specific factors to explain expected payoffs of financial assets, see Fama and French (1992, 1993, 2015). Given several stylized facts about asset returns, such as a stationary auto-regressive pattern, strong time-varying cross-section correlations between series, and clusters of volatility common to all series, more flexible and complex model structures may be better suited for this purpose. In the literature, several dynamic factor models, with different long and short-run dynamics for returns, are shown to be useful in capturing such data properties, see Ng et al. (1992), Quintana et al. (1995), Aguilar and West (2000) and Han (2006) among several others. In a vector autoregressive model the issue of stationarity of the time series strongly influences long-run predictability. In order to capture different stylized facts, we specify several combinations of dynamic models. These models are components of an extended version of a factor-augmented vector autoregressive model, see Bernanke et al. (2005) and Stock and Watson (2005), which includes stochastic volatility and is denoted by FAVAR-SV.

Standard portfolio analysis compares realized returns from different portfolio strategies and assesses the performance. But predicted returns using dynamic models do not lead directly to a practical policy tool for investors, that is, to a decision which portfolio strategy to follow. Common practice is instead to compare realized returns from different portfolio strategies and select the best performing one, see e.g. Aguilar and West (2000). Alternatively, it is possible to incorporate a specific portfolio strategy in the model, but this typically requires a specific model-based strategy such as mean-variance optimization, see e.g. Winkler and Barry (1975), and a specific utility function for the investor, see e.g. Aguilar and West (2000).

In this paper we consider a large set of dynamic models and combinations of models with different short and long-run dynamics in direct connection with different portfolio strategies that an investor can follow. The obtained dynamic asset-allocation can be seen as a mixture of alternative models and alternative portfolio strategies. Our strategy approach extends the mixture of the experts analysis (Jacobs et al., 1991, Jordan and Jacobs, 1994, Jordan and Xu, 1995, Peng et al., 1996), by allowing the strategy weights to be dependent between strategies as well as over time and to further allow for strategy incompleteness. This, to the best of our knowledge, novel methodology provides dynamic asset-allocations, where the underlying models, portfolios as well as portfolio strategies are updated at every decision period. We present an extension of the density combination scheme in Casarin et al. (2016) in order to obtain these time-varying model and portfolio strategy combinations. Using this scheme, we combine alternative models and portfolio strategies in a fully Bayesian setting, where predictive distributions of each model and strategy outcome affects the weights of the strategies and models. In return, the trading strategy, or the policy recommendation to the investor, includes the uncertainty in the strategy and model outcome.

The flexible model and strategy combination structure of the density combinations is estimated using a non-linear and non-Gaussian state space model. This brings a challenge in terms of the estimation robustness and length of computing time, particularly in case of a large number of stocks, a large number of models and strategies. In order to mitigate computing time, we introduce a novel non-linear and non-Gaussian filter: the MitISEM Filter (M-Filter) that is embedded in the density combination procedure. The M-Filter is based on the MitISEM procedure recently proposed by Hoogerheide et al. (2012) and developed in Baştürk et al. (2016). Through a set of simulation studies, we show that the proposed filter is an improvement in terms of the approximation capabilities and computing time compared to other non-linear and non-Gaussian filters such as the Bootstrap Particle Filter (PF) of Gordon et al. (1993) and the Auxiliary Particle Filter (APF) of Pitt and Shephard (1999).

We investigate the performance of the model and portfolio strategy combination method and the M-Filter using US industry portfolios between 1926M7 and 2015M6. Our results show that time-varying combinations of flexible models in the FAVAR-SV class and two momentum strategies lead to better return and risk features than very simple and very complex models. More specifically, the proposed model combinations help to improve return features like mean returns and Sharpe ratios while combinations of two strategies help to reduce risk features like volatility and large loss. The latter result indicates that complete densities provide useful information for risk. We emphasize that our results are conditional upon our data set, US industrial portfolios over the period 1926M7 and 2015M6, as well as our model and strategy set.

The contents of this paper are structured as follows: Section 2 introduces the dynamic models used for US industry returns. Section 3 describes the combined model and portfolio strategies. Section 4 summarizes the density combination scheme and introduces the M-Filter. The approximation and speed performances properties of this novel filter in combining models and portfolio strategies is shown through a set of simulation studies. Section 5 contains the empirical application with 10 US industrial portfolios. Section 6 concludes.

2 Stylized facts about US industrial portfolios and dynamic model structures

In this section we summarize stylized facts about the returns of ten US industry portfolios between 1926M7 and 2015M6 and further key features of dynamic models that will be used to model these returns. Figure 1(a) presents monthly returns of the industry portfolios labelled as 'non-durables', 'durables', 'manufacturing', 'energy', 'hi-tech', 'telecom', 'shops', 'health', 'utilities' and the final category 'others'. The returns of each industry are constructed by equally weighting all stock returns in the specific industry.[1]
We present a descriptive analysis of the correlation between series, co-movements between series, and the change of these over time. In Figure 1(b) the major canonical correlation coefficients between the pairs of series are given. In Figure 1(c) the percentage of variation in the series explained by the first four principal components is shown. Principle components and correlation calculations are based on moving windows with 240 monthly observations.

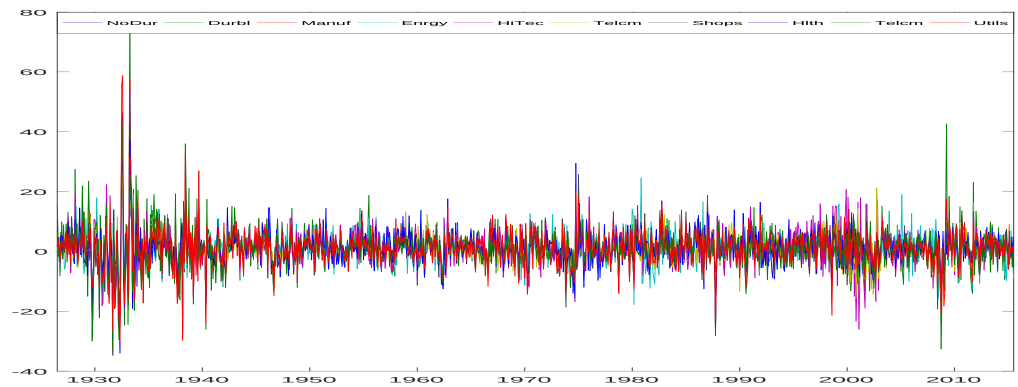
Figures 1(a)–1(c) suggest at least four stylized facts: a stationary auto-regressive time-series pattern for all the series; a strong cross-section correlation between returns with a time-varying pattern; clear volatility clustering, common to all the series; total series variation being well explained by one to four principle components, yet with the explanatory power, hence the number of common factors for the series, being time-varying. Given these data features, we consider different models and combination of models with alternative short and long-run dynamics and distributions. All models considered are members of the following general class of models:

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

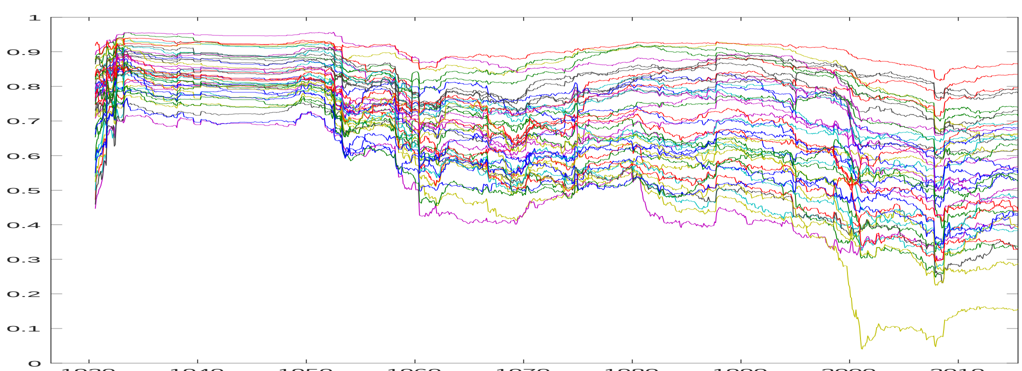
In equation (1), the dependent variable $\mathbf{y}_t = (y_{1t}, \ldots, y_{Nt})'$ is the $N \times 1$ vector of industrial portfolio returns, where y_{it} denotes the return from industry i at time t and the time series runs from $t = 1, \ldots, T$. The $P \times 1$ vector of predetermined variables $\mathbf{x}_t$ can contain explanatory variables or lagged dependent variables. The $K \times 1$ vector $\mathbf{f}_t$ contains unobservable factors with possibly L lags, where ϕ_j for $j = 1, \ldots L$ is a $K \times K$ matrix of autoregressive coefficients. Λ is an $N \times K$ matrix of factor loadings. Finally, $\boldsymbol{\varepsilon}_t$ is an $N \times 1$ vector of idiosyncratic disturbances distributed as $\mathrm{N}(0, \Sigma_t)$ where the time varying variance-covariance matrix may be constant as a special case and $\boldsymbol{\eta}_t$ is a $K \times 1$ vector of latent disturbances distributed as $\mathrm{N}(0, \mathrm{Q}_t)$ where also the time varying variance-covariance matrix may be constant as a special case. Different short and long-run dynamic behaviour of the model in equation (1) are obtained by specifying different restriction on: predeter-

[1] The data are retrieved from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french on 24/10/2015.

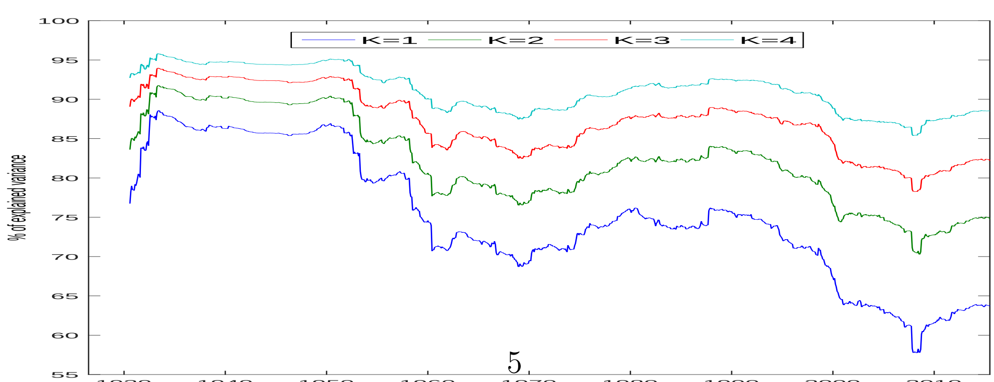
Figure 1: *Monthly percentage returns, explained variation from principle components and canonical correlations 10 US industry portfolios. Monthly return of the industry portfolios Figure 1(a), principle components in Figure 1(b) and correlation calculations in Figure 1(c) are based on moving windows with 240 monthly observations. We use first 50 observations as the initial sample and calculate expanding windows until observation 240.*



(a) Monthly percentage returns



(b) Canonical correlations between 45 pairs



(c) Percentage of explained variation by PCA mined variables $\mathbf{x}_t$, long and short-run lag structure, idiosyncratic disturbances and latent disturbances. All the models considered are reported in Table 1.

Table 1: *Sets of models starting from simple structures and leading to more complex ones.*

Abbreviation	*Full description*
SV	Basic stochastic volatility model
VAR-N	VAR with one lag and normal distributed errors
VAR-SV	VAR with one lag with stochastic volatility in errors
DFM-N	DFM with normal distributed idiosyncratic errors
DFM-SV	DFM with stochastic volatility in idiosyncratic errors
DFM-SV2	DFM with stochastic volatility in the idiosyncratic and in the latent errors
FAVAR-SV	Factor Augmented VAR with stochastic volatility in the idiosyncratic errors
FAVAR-SV2	Factor Augmented VAR with stochastic volatility in the idiosyncratic and in the latent errors

Different short and long-run dynamic behaviour of the model in equation (1) is obtained by specifying different assumptions regarding the predetermined variables $\mathbf{x}_t$, long and short-run lag structure, the idiosyncratic disturbances and the latent disturbances.

The most basic factor model assumes $\boldsymbol{\beta} = 0_{(N\times P)}$, a normal distribution for the idiosyncratic and latent disturbances with time-invariant variance-covariance matrices. We denote this model by DFM, and note that several features of the return data are not well modelled. Another basic model is obtained by letting $\Lambda = 0_{(N\times K)}$ and defining $\mathbf{x}_t$ as the lagged dependent variable. This gives the vector-autoregressive model with a normal distribution for the idiosyncratic disturbances with a time-invariant variance-covariance matrix, denoted by VAR.

The second subclass of models has a stochastic volatility component. When this is specified only in the idiosyncratic disturbances, we denote this with DFM-SV or VAR-SV. When stochastic volatility components are modelled for the idiosyncratic and the latent disturbances, we denote the model by DFM-SV2.

The third group of models is the general model with dynamic factors and lagged dependent variables, and further stochastic volatility in the idiosyncratic and latent disturbances. Our general class of models is an extended version of a factor-augmented VAR model, see Bernanke et al. (2005) and Stock and Watson (2005), which we denote by FAVAR-SV and FAVAR-SV2. We provide details on the specification of the models in Appendix A together with their prior specification and Bayesian estimation procedures.

In our empirical analysis, reported in Section 5, we explore the performance of alternative combinations of models and equity momentum strategies in two steps. We start with the performance of portfolio strategies going from a basic DFM or VAR model to the more complex FAVAR-SV class. As a central issue, we explore the behaviour of time-varying mixtures of basic as well as flexible model structures in combination with two momentum strategies. A reason for this is that the stylized facts of the present section indicate a time-varying pattern in the volatility and cross-correlations.

We end this section with a remark on identification. The general model in equation (1) is not identified without further parameter restrictions. This is clearly seen from the following equality:

$$\mathbf{f}_t\Lambda = \mathbf{f}_t \mathrm{R}\mathrm{R}^{-1}\Lambda,$$

for any $K \times K$ invertible matrix R, which has K^2 free parameters. Hence, at least K^2 restrictions are needed for the model to be identified, see Geweke and Zhou (1996), Lopes and West (2004) and more recently Bai and Peng (2015). We also make use of the restriction of a diagonal covariance matrix. In all models, we follow the identification scheme in Lopes and West (2004).

3 Combining portfolio models and strategies

Conditional upon available information on investment opportunities, portfolio analysis compares realized returns from different strategies and assesses their performance. Econometric models, such as those presented in Section 2, can be used to yield valuable information and result in accurate predictive densities of the dependent variables. Such econometric model forecasts can then serve as input for a portfolio strategy that the investor wants to consider, but this incorporation is not straightforward. Modeling portfolio strategies jointly with return distributions typically requires a strategy such as mean-variance optimization, see e.g. Winkler and Barry (1975) and/or a specific utility or loss function for the investor, see e.g. Aguilar and West (2000).

A novel contribution of this paper is to connect portfolio strategy decisions directly with model comparison and model combination, without the need to specify a loss or utility function for the investor. Our approach is also different from a conventional model combination approach since in our case the main policy decision of an investor, which is deciding on portfolio weights based on alternative investment strategies and underlying models, is directly connected to the model combination. The resulting model and investment strategy combination is a mixture of alternative models and mixture of alternative portfolio strategies that an investor may want to follow.

Several different portfolio strategies are proposed in the literature. Some of these strategies, such as the standard momentum strategy, are not based on a model and are hence hard to incorporate in modeling. Recently, residual momentum strategies are shown to perform well, see e.g. Blitz et al. (2011). Those strategies are obtained using the residuals of a specific model. Stocks with unexpected (surprise) returns in last P periods are shown to perform better than the remaining returns. A residual momentum strategy in practice sorts the last P residuals from the model, e.g. the model $\mathcal{M}_m$ defined by (1) using specific assumptions, and invests in the stocks which have high estimated residuals, and goes short in the stocks which have low residuals over the period P.

We consider two equity momentum strategies based on the fitted return distributions of industry returns from a specific model. In the first strategy, denoted by *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. I.e. the investment decision is based on the model implication directly. In the second strategy, denoted by *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.

In the empirical applications, we apply several model and portfolio strategy combinations in order to obtain profitable 'industry momentum portfolios'. The obtained model and R.M. strategies are similar to Moskowitz and Grinblatt (1999), where the investor decides on the portfolio weights given to each industry. Our approach to form industry portfolios is, however, different from these authors since we consider a large set of models and model combinations and we use a Bayesian approach. Our methodology provides a fully Bayesian framework to account for model uncertainty and strategy uncertainty. From the posterior draws of model parameters, it is straightforward to construct predictive return densities from each strategy. These 'density forecasts' account for parameter uncertainty in the models as well as in the strategy choice. We illustrate this point in detail below.

Constructing a model based portfolio strategy: For a given model $\mathcal{M}_m$ for $m = 1, \ldots, M$, we define a portfolio strategy $\mathcal{S}_s$ for $s = 1, \ldots, S$. For N industries, industry portfolio weights at a portfolio decision time t depends on the strategy and the underlying model, since all portfolio strategies we consider are model-dependent. We denote these weights by the $N \times 1$ vector $\omega_{t,s,m}$ to indicate that these weights are strategy and model dependent.

For calculating model and residual momentum strategies, we use a common measure of the cumulative raw return for the past P observations on the asset to decide on the portfolio weights. In the empirical application to monthly data, we set $P = 12$ as in Jegadeesh and Titman (1993) and Fama and French (1996). For the portfolio decisions, we skip the most recent observation as in standard momentum literature, see e.g. Asness et al. (2013). Then the constructed portfolio is held for P months, after which a new model estimation and portfolio construction is made. We note that increasing the momentum period and the corresponding portfolio holding period mitigates the transaction costs. The realized return of such a portfolio can be calculated as follows:

$$r_{t+P}^{\text{real}} = \iota \cdot \mathbf{y}_{t+1:t+P} \cdot \omega_{t,s,m}, \tag{2}$$

where ι is a $1 \times P$ vector of ones, and $\mathbf{y}_{t+1:t+P} = (\mathbf{y}_{t+1} \ldots \mathbf{y}_{t+P})'$ is the $P \times N$ matrix of realized returns for each stock in the portfolio, after the skip period. In other words, r_{t+P}^{real} is the total return during the holding period.
Different models for returns and different portfolio strategies imply different portfolio weights, $\omega_{t,s,m}$. The portfolio strategies we consider define portfolio weights, $\omega_{t,s,m}$, as a deterministic function of the model, investment strategy and past data points, as explained in detail below.

Model and strategy uncertainty in realized returns: When the portfolio strategy is explicitly based on a model $\mathcal{M}_m$, we show that model and parameter uncertainty can be taken into account in a straightforward way in calculating the realized returns from a strategy. First, when the portfolio construction is based on a specific model $\mathcal{M}_m$, Bayesian inference of the model provides posterior distributions of model parameters together with the weights of the portfolio in (2). Consider a deterministic portfolio strategy $\mathcal{S}_s$ for model $\mathcal{M}_m$, past data points $\mathbf{y}_{t-P+1:t}$ and residuals $\boldsymbol{\varepsilon}_{t-P+1:t}$:

$$\omega_{t,s,m} = g_s(\mathbf{y}_{t-P+1:t}, \boldsymbol{\varepsilon}_{t-P+1:t}^{(m)}), \tag{3}$$

where $\boldsymbol{\varepsilon}_{t-P+1:t} = (\boldsymbol{\varepsilon}_{t-P+1}, \ldots, \boldsymbol{\varepsilon}_t)'$ are the residuals of the model in (1), given by $\omega_{t,s,m} = (\omega_{t,s,m,1}, \ldots, \omega_{t,s,m,N})'$ are the portfolio weights, and $g_s(\cdot)$ is a deterministic function defined by the portfolio strategy. In practice, we obtain posterior draws from these residuals via MCMC, hence draws from the deterministic weights are obtained. Given D posterior draws $\boldsymbol{\varepsilon}_{t-P+1:t}^{(m,d)}$ for $d = 1, \ldots, D$, we obtain draws from the weight distribution:

$$\omega_{t,s,m}^{(d)} = g_s(\mathbf{y}_{t-P+1:t}, \boldsymbol{\varepsilon}_{t-P+1:t}^{(m,d)}). \tag{4}$$

Using (4), realized returns from a specific model and investment strategy in (2) also have a posterior distribution. Posterior draws from this distribution are obtained as follows:

$$r_{t+P}^{\text{real}(d)} = \iota \cdot \mathbf{y}_{t+1:t+P} \cdot \omega_{t,s,m}^{(d)}, \tag{5}$$

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

Model and parameter uncertainty in predicted returns: Similar to the case of realized returns, predicted returns from each model and strategy can also be calculated using the posterior parameter draws in a straightforward way. These *predicted returns from* *strategies and models* constitute the basis of our Bayesian model and strategy combination approach. Specifically, our model combination is based on the one period ahead predictive densities for the 'skip period' in portfolio strategies. This one period ahead predictions of returns can be calculated as follows:

$$\tilde{r}_{t+1}^{(d)} = \mathbf{y}_{t+1}^{(m,d)} \cdot \omega_{t,s,m}^{(d)}, \tag{6}$$

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

We next summarise how the weight function in (3) is defined in different portfolio strategies.

3.1 Standard momentum strategy

We first summarise one of the most common portfolio strategies, standard momentum, which constitutes the baselines for our model comparison. We note that this strategy is not based on a specific model, therefore the distribution of the weights in (3) and the distribution of the realized returns in (2) have a point mass. In addition, the is no underlying model to obtain 'forecast' of returns, $\mathbf{y}_{t+1}^{(m,d)}$, hence (6) cannot be calculated for this strategy.

For the standard momentum strategy, at each portfolio decision time t^p, past performance of the returns are assessed based on the cumulative returns during the past P periods. We consider such a strategy that invests in the top 10% 'winner' stocks and goes short in the bottom 10% 'loser' stocks. Winner and loser stocks are determined according to the past performances. For portfolio weight of each stock is calculated as follows:

$$\omega_{t,s,m,n} = \omega_{t,n} = \begin{cases} 1 & \text{if } \bar{\mathbf{y}}_{n,t} \geq \text{quant}(\bar{\mathbf{y}}_t, 0.90) \\ -1 & \text{if } \bar{\mathbf{y}}_{n,t} \leq \text{quant}(\bar{\mathbf{y}}_t, 0.10) \\ 0 & \text{otherwise} \end{cases}, \tag{7}$$

where $\bar{\mathbf{y}}_{n,t} = \sum_{p=1}^{P} \mathbf{y}_{n,t-p}$ and $\bar{\mathbf{y}}_t = (\bar{\mathbf{y}}_{1,t}, \ldots, \bar{\mathbf{y}}_{N,t})'$ is the set of cumulative returns during the momentum period, $\mathbf{y}_t = (\mathbf{y}_{1,t}, \ldots, \mathbf{y}_{N,t})'$ and $\text{quant}(\mathbf{x}, p)$ denotes the $100 \times p$ percent quantile of the elements of vector $\mathbf{x}$. The breakpoints of momentum, 90% and 10%, can be adjusted for different portfolio strategies.

3.2 Density combination

A density combination approach usually consists of a convolution of three classes of densities: a model combination density, a weight density and a density of the predictors of many models. Casarin et al. (2016) provide a representation of this approach as a large finite mixture of convolutions of densities of different models which generalizes the mixture of experts approach experts analysis from (Jacobs et al., 1991, Jordan and Jacobs, 1994, Jordan and Xu, 1995, Peng et al., 1996) by allowing the weights to be dependent over time and between models. For the time-varying weights a learning mechanism is specified and the approach also allows for incompleteness of the model set and the set of strategies. We note that for efficient computational purposes use is made of GPU and parallel computing.

For convenience, we present a brief summary of the combination of densities using one economic variable. The *conditional predictive probability* of an economic variable of interest y_t, given a set of K predicted variables, $\tilde{\mathbf{y}}_t = (\tilde{y}_{1t}, \dots, \tilde{y}_{Kt})'$, is specified as a *discrete mixture of conditional predictive probabilities* of y_t given $\tilde{y}_{kt}$ using the following formula:

$$f(y_t|\tilde{\mathbf{y}}_t, \mathrm{I}_{\mathbf{K}}) = \sum_{k=1}^{K} w_{kt} f(y_t|\tilde{y}_{kt}, I_k), \tag{8}$$

where w_{kt}, $k = 1, \dots, K$ are the weights of each individual models or experts, I_k with $k = 1, \dots, K$ is the information set of model k and $\mathrm{I}_{\mathbf{K}}$ is the joint information set for all models.

Given equation (8) we can derive, under standard regularity conditions (a Markov process for the weights and conditional independence of the density of y_t given the past information), that the *marginal* predictive density of y_t has the following discrete/continuous representation:

$$f(y_t|\mathrm{I}_{\mathbf{K}}) = \sum_{k=1}^{K} w_{kt} \int_{\mathbb{R}} f(y_t|\tilde{y}_{kt}, I_k) f(\tilde{y}_{kt}|I_k) d\tilde{y}_{kt}. \tag{9}$$

In Casarin et al. (2016), it is specified that the weights have a logistic dynamics described by the additive logistic transform

$$w_{kt} = \exp\{x_{kt}\} / \sum_{k=1}^{K} \exp\{x_{kt}\}, labeleq : WeightsMix \tag{10}$$

where $x_{it} \in \mathbb{R}$ is a latent Gaussian process. Here the past predictive performance or different economic information can be specified. A learning mechanism can also be added to the weight dynamics.

In order to evaluate this density combination scheme, Casarin et al. (2016) show that it can we written as a non-linear and non-Gaussian combinational scheme that yields a forecast density for the observable variables, conditional on the predictors and on the combination weights. This representation is quite general, but requires the use of a sequential Monte Carlo algorithm in order to numerically evaluate results.

We define $\mathbf{\Gamma}_t = \mathrm{vec}(\mathrm{W}_t)$ as the vector of model weights associated with $\tilde{\mathbf{y}}_t$ and $\boldsymbol{\theta} \in \Theta$ the parameter vector of the combination model. In addition, we define the augmented state vector $\boldsymbol{\alpha}_t = (\boldsymbol{\Gamma}_t, \boldsymbol{\theta}_t)$ where $\boldsymbol{\theta}_t = \boldsymbol{\theta}$, $\forall t$. Using these definitions, the distributional state space form of the forecast model is:

$$\begin{aligned} \mathbf{y}_t &\sim p(\mathbf{y}_t|\boldsymbol{\alpha}_t, \tilde{\mathbf{y}}_t) && \text{(measurement density)}, \\ \boldsymbol{\alpha}_t &\sim p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}, \mathbf{y}_{1:t-1}, \tilde{\mathbf{y}}_{1:t-1}) && \text{(transition density)}, \\ \boldsymbol{\alpha}_0 &\sim p(\boldsymbol{\alpha}_0) && \text{(initial density)}. \end{aligned} \tag{11}$$

The state predictive and filtering densities conditional on the predictive variables $\tilde{\mathbf{y}}_{1:t}$ are:

$$\begin{aligned} p(\boldsymbol{\alpha}_{t+1}|\mathbf{y}_{1:t}, \tilde{\mathbf{y}}_{1:t}) &= \int p(\boldsymbol{\alpha}_{t+1}|\boldsymbol{\alpha}_t, \mathbf{y}_{1:t}, \tilde{\mathbf{y}}_{1:t}) \\ &\qquad p(\boldsymbol{\alpha}_t|\mathbf{y}_{1:t}, \tilde{\mathbf{y}}_{1:t}) d\boldsymbol{\alpha}_t, \\ p(\boldsymbol{\alpha}_{t+1}|\mathbf{y}_{1:t+1}, \tilde{\mathbf{y}}_{1:t+1}) &= \frac{p(\mathbf{y}_{t+1}|\boldsymbol{\alpha}_{t+1}, \tilde{\mathbf{y}}_{t+1}) p(\boldsymbol{\alpha}_{t+1}|\mathbf{y}_{1:t}, \tilde{\mathbf{y}}_{1:t})}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}, \tilde{\mathbf{y}}_{1:t})}, \end{aligned}$$

respectively, which represent the optimal non-linear filter, see Doucet et al. (2001). The marginal predictive density of the observable variables is then:

$$p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}) = \int p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}, \tilde{\mathbf{y}}_{t+1}) p(\tilde{\mathbf{y}}_{t+1}|\mathbf{y}_{1:t}) d\tilde{\mathbf{y}}_{t+1},$$

where $p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}, \tilde{\mathbf{y}}_{t+1})$ is defined as:

$$\int p(\mathbf{y}_{t+1}|\boldsymbol{\alpha}_{t+1}, \tilde{\mathbf{y}}_{t+1}) p(\boldsymbol{\alpha}_{t+1}|\mathbf{y}_{1:t}, \tilde{\mathbf{y}}_{1:t}) p(\tilde{\mathbf{y}}_{1:t}|\mathbf{y}_{1:t-1}) d\boldsymbol{\alpha}_{t+1} d\tilde{\mathbf{y}}_{1:t},$$

and represents the conditional predictive density of the observable given the past values of the observable and of the predictors.

3.3 Density combinations for mixtures of models and investment strategies

In the previous section we outlined the general predictive density combination scheme in Casarin et al. (2016), where model weights are based on the predicted distributions of observed data $\mathbf{y}_t$. Such a combination of predicted model returns, e.g. predicted returns for each stock in $\mathbf{y}_t$, do not lead directly to a practical policy tool for investors, that is, to a decision about which portfolio strategy to follow.

We propose to combine models but also the investment strategies, where the combination scheme takes into account the predictive distributions of the returns of a specific model and investment strategy, given in (6). The corresponding combination scheme, extending

(9), is as follows:

$$f(r_t|\mathrm{I}_{\mathbf{K}}) = \sum_{k=1}^{K} w_{kt} \int_{\mathbb{R}} f(r_t|\tilde{r}_{kt}, I_k) f(\tilde{r}_{kt}|I_k) d\tilde{r}_{kt}, \tag{12}$$

where $k = 1, \ldots, K$ denotes a specific model and investment strategy combination, $(\mathcal{M}_m, \mathcal{S}_s)$ in Sections 2 and 3. In addition, $\tilde{r}_{kt}$ is a scalar, one period ahead return forecast from model and investment strategy combination k at time t. Posterior draws of $\tilde{r}_{kt}$ are obtained from the posterior parameter draws and the investment rule as shown in (6).

Our extension of the combination scheme has a major difference compared to the standard model combination in (9): Actual return, r_t on the left hand side of (12) is not an observed variable. Instead, the objective of the combination scheme is to maximize realized return r_t. Such 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' or 'full information return' in the forecast period using a 'skip month' in the investments. The optimal return r_t is defined as the maximum possible return given the information during the skip month t, under the constraint that portfolio weights sum up to 0. These returns correspond to a strategy that goes long in the asset with the highest return in the skip month, and goes short in the asset with the lowest return in the skip month. We note the above notational difference between the combination schemes in (9) and in (12). In the remainder of this paper we follow the former convention to facilitate a direct comparison of the proposed filtering method with the one in Casarin et al. (2016).

4 Density combinations and the M-Filter

The proposed dynamic asset-allocation can be seen as a mixture of alternative models and alternative portfolio strategies with sequential updating. In order to implement this approach, we make use of the general density combination approach developed in Billio et al. (2013), Casarin et al. (2015) and recently Casarin et al. (2016). We extend this approach that is aimed at forecasting to a mixture of forecasting and strategy combinations. In this section, we start with a brief summary of the existing density combination approach. Next, in order to improve the computational efficiency of the procedure, we present the M-Filter that is used in the combination methodology. Finally, we present details of the model and portfolio strategy combinations using both density combinations and the M-Filter.

4.1 The M-Filter

The M-Filter is based on the MitISEM procedure recently proposed by Hoogerheide et al. (2012) and developed in Baştürk et al. (2016). The state space model of equations (8) and (**??**) is given by the following system:

$$\begin{aligned} \mathbf{y}_t &= m_t(\boldsymbol{\alpha}_t, \tilde{\mathbf{y}}_t, \boldsymbol{\varepsilon}_t), \\ \boldsymbol{\alpha}_t &= h_t(\boldsymbol{\alpha}_{t-1}, \boldsymbol{\eta}_t). \end{aligned} \tag{13}$$

The functions m_t and h_t are possibly non-linear, $\mathbf{y}_t$ are the observations, $\tilde{\mathbf{y}}_t$ are the predictive densities, $\boldsymbol{\alpha}_t$ are the state variables and $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_t$ are mutually independent errors.

Using these definitions, the distributional state space form of the forecast model is:

$$\begin{aligned} \mathbf{y}_t &\sim p(\mathbf{y}_t|\boldsymbol{\alpha}_t, \tilde{\mathbf{y}}_t) && \text{(measurement density)}, \\ \boldsymbol{\alpha}_t &\sim p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}, \mathbf{y}_{1:t-1}, \tilde{\mathbf{y}}_{1:t-1}) && \text{(transition density)}, \\ \boldsymbol{\alpha}_0 &\sim p(\boldsymbol{\alpha}_0) && \text{(initial density)}. \end{aligned} \tag{14}$$

The object of interest it the join conditional distribution defined as:

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

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

Conditionally on the estimated parameters $\hat{\theta}$ the filtering proceeds using the particle filter as follows:

1) **Initialization**. Draw the initial particles from the distribution $\boldsymbol{\alpha}_0^{(j)} \sim p(\boldsymbol{\alpha}_0)$ and set the weights $\mathrm{W}_0^{(j)} = 1/M$ for $j = 1, \ldots, M$.

2) **Recursion**. For $t = 1, \ldots, T$ draw $\tilde{\boldsymbol{\alpha}}_t^{(j)}$ from the density $g_t(\tilde{\boldsymbol{\alpha}}_t^{(j)}|\boldsymbol{\alpha}_{t-1}^{(j)}, \hat{\theta})$, compute the importance weights

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

and normalise them

$$w_t^{(j)} = \frac{\tilde{w}_t^{(j)}}{\sum_{j=1}^M \tilde{w}_t^{(j)}}.$$

3) **Estimation.** The approximation to $\mathrm{E}[h_t(\tilde{\boldsymbol{\alpha}}_t)|\mathbf{y}_{1:t},\theta]$ is given by:

$$\tilde{h}_{t,M} = \sum_{j=1}^{M} w_t^{(j)} h_t(\tilde{\boldsymbol{\alpha}}_t^{(j)}).$$

4) **Selection**. Re-sample the particle via e.g. multinomial resampling.

5) **Likelihood Approximation**. The approximation to the log likelihood is given by:

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

The performnce of any PF algorithm crucially depends on the specification of the proposal density $g_t(\tilde{\boldsymbol{\alpha}}_t^{(j)}|\boldsymbol{\alpha}_{t-1}^{(j)},\hat{\theta})$, and there is a broad literature on how to select it, cf/ Doucet et al. (2001), Liu (2001), Kunsch (2005) and Creal (2012). An inherent property of Sequential Monte Carlo is particle degeneracy, resulting in the weights distribution becoming more and more skewed over time and eventually putting all of its mass in a single particle. The resampling step is necessary to overcome this issue, however it also introduces an additional Monte Carlo variation into the algorithm and slows down the filtering procedure.

Our approach uses the MitISEM algothm of Hoogerheide et al. (2012) to construct the importance density $g_t(\tilde{\boldsymbol{\alpha}}_t^{(j)}|\boldsymbol{\alpha}_{t-1}^{(j)})$ approximating the target density $p(\mathbf{y}_t|\tilde{\boldsymbol{\alpha}}_t^{(j)},\tilde{\mathbf{y}}_t)p(\tilde{\boldsymbol{\alpha}}_t^{(j)}|\boldsymbol{\alpha}_{t-1}^{(j)})$ at each time t. It has two major advantages. First, the MitISEM algorithm is based on mixtures of Student t distributions, which can easily approximate complex non-elliptical posterior distributions, e.g. multimodal and skewed. second, a high quality of the obtained approximation mitigates the problem of particle degeneracy and hence allows us to perform the resmapling step less often.

The proposed M-Filter algorithm, fully described in Appendix B, can be summarized as follows.

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 the candidate density $\tilde{g}_t(\tilde{\boldsymbol{\alpha}}_t^{(j)}|\boldsymbol{\alpha}_{t-1}^{(j)})$ using the MitISEM algorithm, which can be summarize as follows:

   a) **Initialization**: Simulate draws $\tilde{\boldsymbol{\alpha}}_t^{(j)}$ from a 'naive' candidate distribution with density $g_n(\cdot)$ (e.g. a Student-t with $v = 5$ degrees of freedom). Compute the corresponding IS weights:

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

where the target density kernel has the form $p(\mathbf{y}_t|\tilde{\boldsymbol{\alpha}}_t^{(j)}, \tilde{\mathbf{y}}_t)p(\tilde{\boldsymbol{\alpha}}_t^{(j)}|\boldsymbol{\alpha}_{t-1}^{(j)})$, and normalise them to $w_t^{(j)}$.

b) **Adaptation**: Improve the candidate density using the MitISEM algorithm (cf. Appendix B).

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

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

4) **Likelihood Approximation**. The approximation of the log likelihood function is given by:

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

Here we present the Monte Carlo experiments to show the performance of the MF. In all the examples we are interested in the estimation of the target function $h_t(\boldsymbol{\alpha}_t^{(j)}) = \boldsymbol{\alpha}_t$ that is the posterior mean of the latent state. We compare four filters, the Kalman filter (KF), the Particle Filter (PF), the Auxiliary Particle filter (APF) and the MF.[2] All Monte Carlo experiments presented in this section are based on I = 100 replications, with T = 100 observations each. For the PF, the APF and the M-Filter we use $M = 50,000$ particles.

The first model we consider is a standard local level model:

$$\begin{aligned} y_t &= \alpha_t + \varepsilon_t \qquad \varepsilon_t \sim \mathrm{N}(0, \sigma_\varepsilon^2), \\ \alpha_t &= \alpha_{t-1} + \eta_t \quad\;\; \eta_t \sim \mathrm{N}(0, \sigma_\eta^2). \end{aligned} \tag{15}$$

This linear and Gaussian model is often use as benchmark for comparing filtering methods. In this case KF provides the sequential state distribution in analytical form and is the optimal filter. In the simulations reported below, we fix the latent state variance at $\sigma_\eta^2 = 0.1$ and we define four different levels for the state variance σ_ε^2, corresponding to four levels of the Noise to Signal Ratio (NtS): 0.1, 0.5, 1 and 2.5. We note that the exact likelihood of the local level model in (15) can be calculated using the KF. We compare the exact likelihood from KF in this model and compare it with the remaining non-linear filters to assess the degree of the bias in the non-linear filters, including the proposed M-Filter.

Table 2 reports the results for the model (15). As expected, the KF filter is the best filter in terms of the minimum mean errors and the smallest computing time. The results of

[2]We note that our proposed filter approach is related to that of Liesenfeld and Richard (2003) and Richard and Zhang (2007). An important difference is our flexible and robust procedure to choose the candidate density. We leave a comparison between the two approaches as a topic for future research.

the non-linear filters, however, are in line with those of the KF. The proposed M-Filter performs similarly to the PF and the APF but has a lower bias in the estimate likelihood especially for smallest NtS ratio of 0.1. In the all cases the computing time is lower then the PF and APF.

The second model we consider is the stochastic volatility model (Kim et al., 1998) specified as follows

$$\begin{aligned} y_t &= e^{(\alpha_t/2)}\varepsilon_t & \varepsilon_t &\sim \mathrm{N}(0, \sigma_\varepsilon^2), \\ \alpha_t &= \mu + \phi\alpha_{t-1} + \eta_t & \eta_t &\sim \mathrm{N}(0, \sigma_\eta^2), \end{aligned} \tag{16}$$

where η_t and ε_t are independent and y_t is the observed series. Due to the non-linear structure of the observation equation the analytical form for filtering and predictive densities do not exist in this model. In the simulations, we fix the autoregressive parameter ϕ at 0.90, 0.95, and 0.98, which are in line with the values found in other studies, see for example Aguilar and West (2000).For each value of ϕ we consider four values of σ_η^2, corresponding to the coefficient of variation (CV) of the volatility $h = \bar{\sigma}^2 \exp(\alpha_t)$ defined as

$$\mathrm{CV} = \frac{\mathrm{Var}(h)}{\mathrm{E}(h)^2} = \exp\left(\frac{\sigma_\eta^2}{1-\phi^2}\right) - 1,$$

taking the values 0.1, 0.5, 1, and 2.5. We note that a high value of the CV indicates the relative strength of the volatility process and low values of CV indicate the volatility is close to constant.

Table 3 reports the results for the model (16). In all the cases the KF is the worst filter due to being a linear Gaussian filter. The M-Filter performs similarly to the PF and the APF in term of the MAE and the MSE. In this model the computational speed is comparable between the three non-linear filters, namely the PF, the APF and the M-Filter.

Finally, we examine a DFM that is a multivariate model given by:

$$\begin{aligned} \mathbf{y}_t &= \Lambda\mathbf{f}_t + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t &\sim \mathrm{N}(0, \Sigma), \\ \mathbf{f}_t &= \phi_1\mathbf{f}_{t-1} + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim \mathrm{N}(0, \mathrm{Q}). \end{aligned} \tag{17}$$

Due to the linear Gaussian model structure in (17) the KF is the optimal filter, hence we use the KF as the benchmark for the non-linear filters. Table 4 reports the results for this model for $I = 100$ Monte Carlo replication, $N = 20$ series and 2, 4, 6 and 10 factors. Obviously, the KF leads to the best results in terms of the speed and accuracy, but the non-linear filter results are in line with those from the KF. The M-Filter performs better then the PF and the APF, with substantially lower bias and variance. The M-Filter also leads to the lowest bias in the estimated likelihood compared to the other non-linear and non-Gaussian filters. The computing time for all the filters increases with the number of factors. In all the cases, however, the M-Filter requires the shortest computing time among all the nonlinear non-Gaussian filters. Based on the set of simulation studies, we conclude that the proposed M-Filter has advantages in terms of reduced errors and computing time compared to the considered nonlinear non-Gaussian filters.

Table 2: *Monte Carlo results for I=100 replications of the linear and Gaussian model of equation (15) with T=100 for the Kalman Filter (KF), the Bootstrap Particle (PF), the Auxiliary Particle Filter (APF) and the MitISEM Filter (M-Filter) with 50,000 particles. The table reports Likelihood Bias (LB) with respect to KF, the Mean Absolute Error as* $MAE = 1/I \sum_{i=1}^{I} abs(\tilde{\alpha}_{t,i} - \alpha_{t,i})$ *relative to the KF, the Means Squared Error defined as* $MSE = 1/I \sum_{i=1}^{I} (\tilde{\alpha}_{t,i} - \alpha_{t,i})^2$ *relative to the KF. The final column reports the computing time in seconds for the four filters.*

NtS		*0.1*			*0.5*		*Time with NtS*	
Model	LB	MAE	MSE	LB	MAE	MSE	0.1	0.5
KF	0.00	1.00	1.00	0.00	1.00	1.00	0.003	0.003
PF	-48.93	1.22	1.48	-19.43	1.26	1.62	33.711	35.549
APF	-13.87	1.00	1.00	-9.56	1.01	1.02	35.542	37.673
M-Filter	-10.40	1.00	1.01	-9.52	1.01	1.02	12.831	12.814
NtS		*1*			*2.5*		*Time with NtS*	
Model	LB	MAE	MSE	LB	MAE	MSE	1	2.5
KF	0.00	1.00	1.00	0.00	1.00	1.00	0.003	0.003
PF	-37.85	1.31	1.71	-21.16	1.43	2.04	35.219	34.531
APF	-10.43	1.00	1.00	-9.05	1.00	1.00	37.295	35.721
M-Filter	-10.18	1.01	1.01	-9.39	1.00	1.01	12.668	12.126

5 Empirical application using US industrial portfolios, 1926-2015

In this section we apply different dynamic models, investment strategies and combinations of models and strategies to monthly returns from ten US industry portfolios between 1926M7 and 2015M6. We note that each member of the DFM class is estimated with a different correlation structure, defined through the number of factors and the number of factor lags. In addition, for all models we construct portfolios based on two investment strategies, M.M. and the R.M. strategy, presented in Section 3.

We start to analyze the performance of individual models combined with a particular strategy on returns and risk, using as indicators: means, volatilities, Sharpe Ratios and the largest loss during the investment period. In addition, we compare the results of the proposed models and investment strategies with a baseline standard momentum strategy as

Table 3: *Monte Carlo results for I=100 replications of the stochastic volatility model* (16) *with T=100for the Kalman Filter (KF), Bootstrap Particle Filter (PF),the Auxiliary Particle Filter (APF) and the MitISEM Filter (M-Filter), with 50,000 particles. The table reports Likelihood Bias (LB) with respect to KF, the Mean Absolute Error defined as* $MAE = 1/I \sum_{i=1}^{I} abs(\tilde{\alpha}_{t,i} - \alpha_{t,i})$ *relative to the KF,the Mean Squared Error defined as* $MSE = 1/I \sum_{i=1}^{I} (\tilde{\alpha}_{t,i} - \alpha_{t,i})^2$ *relative to the KF. The final column reports the computing time in seconds for the four filters.*

CV	*0.1*		*0.5*		*Time*	
Model	MAE	Var	MAE	Var	0.1	0.5
KF	1.00	1.00	1.00	1.00	0.003	0.003
PF	0.24	0.10	0.31	0.12	13.817	13.989
APF	0.25	0.10	0.31	0.13	14.583	14.666
M-Filter	0.26	0.10	0.31	0.14	14.145	12.670
CV	*0.1*		*0.5*		*Time*	
Model	MAE	Var	MAE	Var	1	2.5
KF	1.00	1.00	1.00	1.00	0.003	0.003
PF	0.32	0.12	0.29	0.11	13.982	13.876
APF	0.31	0.13	0.29	0.11	14.612	14.704
M-Filter	0.30	0.13	0.28	0.11	13.541	12.963

Table 4: *Monte Carlo results for* $I = 100$ *replications of the DFM* (17) *with* $T = 100$ *and* $N = 20$ *for the Kalman Filter (KF), Bootstrap Particle Filter (PF),the Auxiliary Particle Filter (APF) and the MitISEM Filter (M-Filter), with 50,000 particles. The table reports Likelihood Bias (LB) with respect to KF, the Mean Absolute Error defined as* $MAE = 1/I \sum_{i=1}^{I} abs(\tilde{\alpha}_{t,i} - \alpha_{t,i})$ *relative to the KF,the Mean Squared Error defined as* $MSE = 1/I \sum_{i=1}^{I} (\tilde{\alpha}_{t,i} - \alpha_{t,i})^2$ *relative to the KF. The final column reports the computing time in seconds for the four filters in case of 2, 4, 6 and 10 latent factors.*

Factors		*2*			*4*		*Time*	
Model	LB	MAE	MSE	LB	MAE	MSE	2	4
KF	0	1	1	0	1	1	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
M-Filter	-23.23	1.01	1.02	-23.39	1.00	1.01	106.330	138.178
Factors		*6*			*10*		*Time*	
Model	LB	MAE	MSE	LB	MAE	MSE	6	10
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
M-Filter	-16.97	1.03	1.03	-112.68	1.02	1.03	213.200	402.820

presented in Jegadeesh and Titman (1993), Chan et al. (1996) and Jegadeesh and Titman (2001).

The central issue that we address is to explore possible time-variation in the performance of combinations of models and investment strategies. Here, the time-varying weights in the combination scheme are used to identify and estimate the effect that the stylized data features, listed in Section 2, may have on model forecasts and equity momentum results.

We note that the investment time-line for all model and strategy choices is specified as follows: Investment decisions are made in July every year, where all models are estimated at the investment decision times using 240-monthly data windows. A portfolio is held for 12 months after a skip period of one month, namely August. All models are re-estimated annually at the investment decision month, July, leading to 69 moving window estimates. For the two investment strategies calculations are done using model-based results from the past 12 months, see Jegadeesh and Titman (1993) and Fama and French (1993). We further note that a skip month in portfolio calculation is often used to remove market microstructure effects, see Asness et al. (2013). In this analysis, we use this 'skip month' to calculate predictive performances of models and investment strategies, as explained in Section 4.

5.1 Return and risk features from combining individual model forecasts and investment strategies

In this subsection we report realized return and risk properties of several model and investment strategies together with those of the standard momentum strategy. For all the models, Bayesian inference is performed based on 5000 posterior draws, after a burn-in of 5000 draws. In total, we estimate 40 combinations of DFM models, 2 VAR models (VAR and VAR-SV), and an SV-model. For each DFM model, we consider 8 different specifications which correspond to 1-4 factors and 1-2 lags for the factor equation. We restrict the dynamics of the VAR-class to one lag. Given 10 data series, already a VAR(1) gives very flexible dynamic patterns (shown in their implied moving averages). For each one of these models, we consider two investment strategies and construct portfolios based on the M.M. strategy and the R.M. strategy. This leans to 86 combinations of models and investment strategy specifications, summarized in Table 5. We consider eight sets of models, starting from simple structures and leading to more complex ones, namely: a VAR-N, SV and DFM-N model with normally distributed disturbances.

Next, a DFM model and a VAR model with stochastic volatility components in idiosyncratic disturbances (DFM-SV and VAR-SV) and a DFM model with stochastic volatility components in idiosyncratic disturbances and in the factor equation (DFM-SV2). Third, a FAVAR model with stochastic volatility components in idiosyncratic errors (FAVAR-SV) and a FAVAR model with stochastic volatility components in idiosyncratic errors and in the factor equation (FAVAR-SV2).

Table 5: *Returns and risk for 10 industry portfolios. The Table reports the mean, volatility (Vol.), Sharpe Ratio (S.R.) and the largest loss (L.L.) for realized returns for all models and strategies in Section 2. The investment strategies are: Model Momentum and Residual Momentum. The standard momentum strategy has mean 0.09, volatility 5.7, Sharpe ratio 0.02 and largest loss -26.2. Bold values indicate an 'equal or better' value compared to standard momentum. K is the number of factors and L is the number of lags.*

		Model Momentum				Residual Momentum			
	(K, L)	Mean	Vol.	S.R.	L.L.	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
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

For each DFM model, we consider 8 different specifications which correspond to 1-4 factors and 1-2 lags for the factor equation, and two investment strategies corresponding to model-momentum and residual momentum. In total, we estimate 40 combinations of DFM models, 2 VAR models (VAR and VAR-SV), and an SV-model. We restrict the dynamics of the VAR-class to the case of one lag. Given 10 data series, a VAR(1) gives already very flexible dynamic patterns (shown in their implied moving averages). For each one of these models, we construct portfolios based on the model-based momentum strategy and a residual momentum strategy. Hence the combination of model and investment strategy specifications leads to 86 results, summarized in Table 5. We report there means, volatilities, Sharpe Ratios and largest loss of each model and strategy.

We first focus on the mean returns in Table 5. The results differ substantially over alternative model and strategy combinations. It can be seen that the M.M. strategy gives poor results for the simple VAR-N and DFM-N models. The complex model structures, DFM-SV2 and FAVAR-SV2, lead to neither better nor worse results than these for the more basic models, DFM-SV and FAVAR-SV, for both momentum strategies. That is, the SV2 component leads mostly to over-fitting, not to better mean returns. A second conclusion is that including the SV component in the VAR, the DFM and the FAVAR models leads to substantially better results for both strategies. It is noteworthy that the choice of the number of factors and lags in the factor models strongly influences thr results in this case. Further, mean returns of these model and strategy combinations are equal (FAVAR-SV) or higher than those of the standard momentum strategy. In summary, there exist clear differences in the results between the two strategies: more complex model structures are good in combination with the M.M., while using simpler models in combination with the R.M. already leads to a good performance on mean returns. Apparently, adopting the latter strategy implies a *learning from past errors* mechanism.

We next compare the volatility of realized returns in Table 5. The differences between realized return volatilities between model and strategy combinations are less pronounced than the differences in mean realized returns. The obtained volatilities from each model and strategy combination are also close to that of the standard momentum strategy. An interesting point is the comparison of model based and R.M. strategies. Given the same model class, M.M. generally leads to a lower volatility compared to residual momentum, but this difference is sensitive for the choice of the number of factors and lags in the factor models.

Since the Sharpe ratios, reported in Table 5, are defined as scaled means and given that the volatility estimates are rather similar over different model structures and strategies, the conclusions about the means listed above hold, for almost all models, also for the Sharpe ratios.

As an indicator of risk, we compare the largest loss from each model and strategy in

Table 5. Simple models and the complex DFM-SV2 and FAVAR-SV2 lead to large losses or do not improve over the other models. Contrary to the positive results on the mean returns, it is of interest to observe that a pure SV model has substantial risk of a loss for the M.M. strategy. A complex model like a FAVAR-SV leads to a very low loss. For all the models, except the SV, the largest loss is substantially lower for the M.M. strategy compared to R.M. Clearly, the choice of a momentum strategy matters substantially for risk of returns.

So far we have focused on summarizing posterior measures. An important advantage of Bayesian inference is delivering complete distributions of the realized returns from the specific model and investment strategy combination. Figure 2 presents the percentiles of the realized returns for three selected model and portfolio strategies, DFM model with $K = 1, L = 1$ and the model momentum, DFM-SV model with model momentum ($K = 4, L = 1$) and DFM-SV model with R.M.($K = 3, L = 2$). First, 99% intervals of returns are relatively tight for all three models.[3] Second, even the worst-performing model and strategy combination, DFM with $K = 1, L = 1$ and M.M., has very high returns in some periods. Similarly, the best-performing strategies, DFM-SV, R.M, with $K = 3, L = 2$ and DFM-SV, M.M., $K = 4, L = 1$ lead to extreme losses in some periods. Apparently, the time-variation in the performances of each model and strategy combination is important and we investigate that in the next subsection.

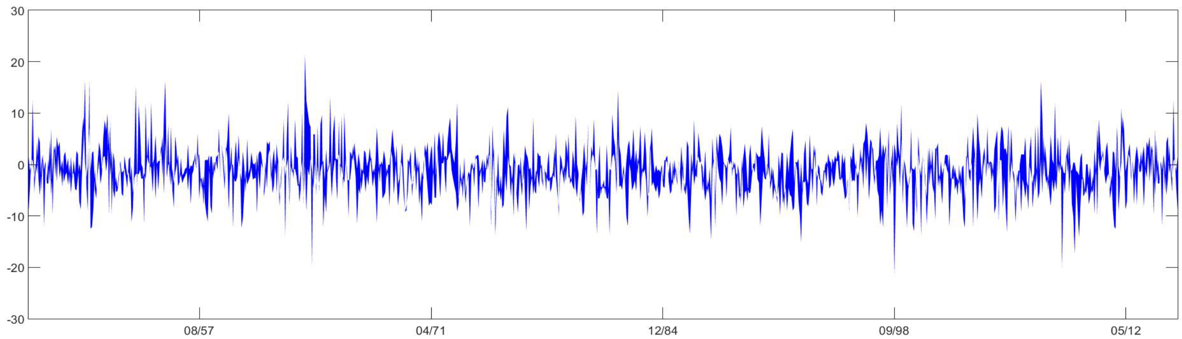
We summarize our conclusions as follows:

- *Diversity of results*: Results for return and risk are sensitive for model choice and strategy.
- *Return and risk features*: Using M.M. in combination with very simple models, like VAR-N and DFM-N that do not fit well, gives poor results. Complex models, like DFM-SV2 and FAVAR-SV2 that overfit, do not improve results compared to the models DFM-SV and FAVAR-SV. R.M. leads to reasonable returns for the simple models, VAR-N, SV and DFM-N. However, it also does not improve the results for the complex DFM-SV2 and FAVAR-SV2 models compared to their simpler cases. Using M.M. gives reasonable risk results for almost all models. There exists a sensitivity in the DFM class for the number of factors and lags. SV has poor risk. Using R.M. gives poor risk results for all models except SV.

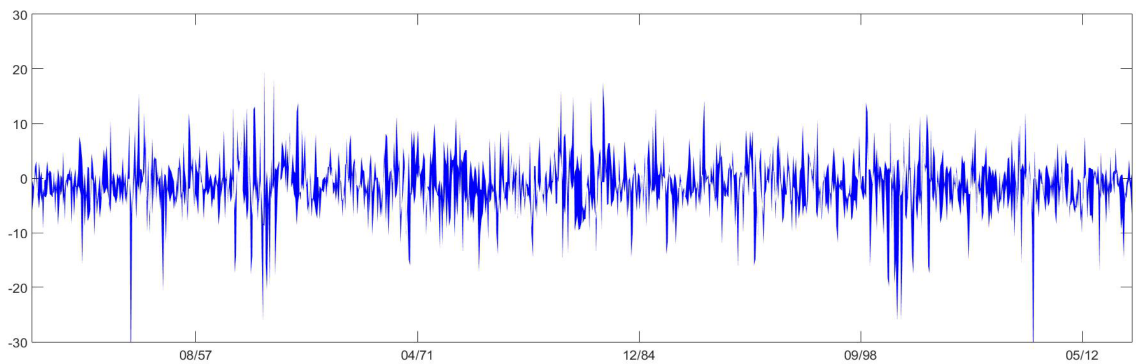
These conclusions lead naturally to our main topic of exploring return and risk features by combining models and strategies and exploring their behaviour over time.

[3]This is a general property for all models we consider. We do not present all the return intervals due to space considerations.

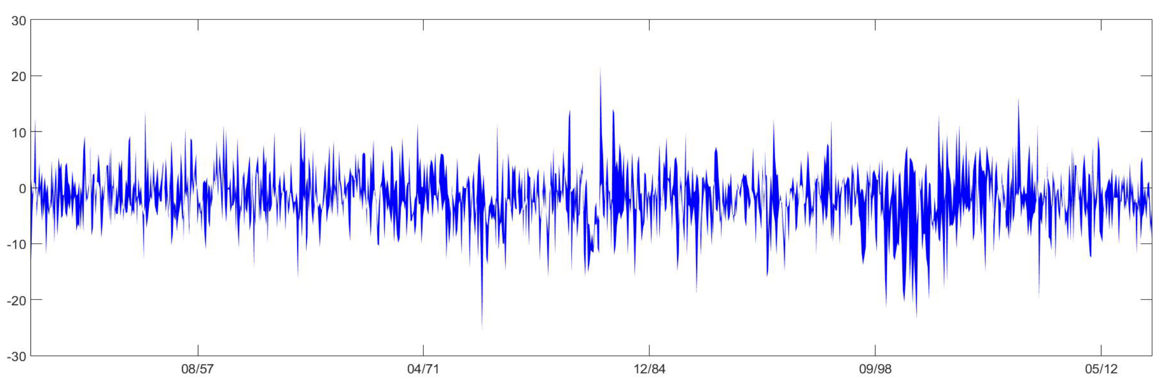
Figure 2: *99% realized return intervals for three selected models. DFM with $K = 1$ latent factor and $L = 1$ lags in Figure 2(a), DFM-SV with $K = 3$ latent factors and $L = 2$ lags in Figure 2(b) and DFM-SV with $K = 4$ latent factors and $L = 1$ lags in Figure 2(c).*



(a) DFM, Model Momentum, K = 1, L = 1



(b) DFM-SV, Residual Momentum, K = 3, L = 2



(c) DFM-SV, Model Momentum, K = 4, L = 1

Table 6: *Returns and risk from mixtures of three basic models and two investment strategies. The top panel shows return and risk from a mixture of three models (VAR-N, SV, DFM-N(4,2)) combined with a mixture of two investment strategies (M.M., R.M.). The bottom panel reports these results for the mixture of two investment strategies combined with each model separately. Standard momentum strategy has mean 0.09, volatility 5.7, Sharpe ratio 0.020 and largest loss -26.2. 90% credible intervals are reported in parentheses.*

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

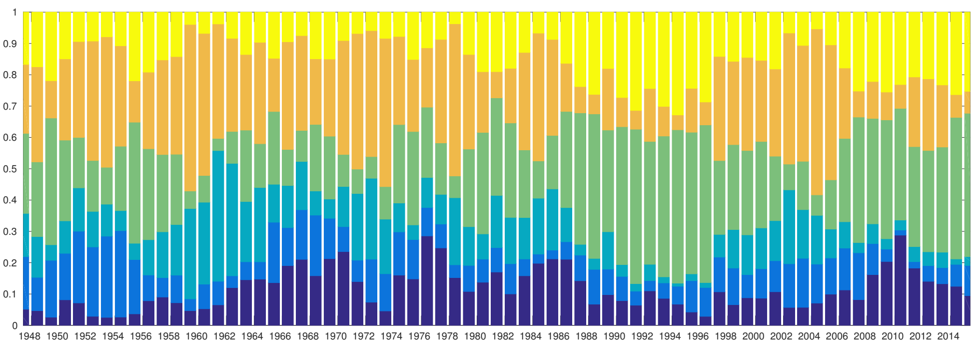
5.2 Combining possible mixtures of models and strategies

We report the time-varying performances of different model and strategy combinations in two stages. First, the three basic model structures, VAR-N, SV and DFM-N(4,2), that constitute together the FAVAR-SV(4,2) class are combined with a mixture of two strategies. Next, we combine two more flexible model structures than the basic ones with a mixture of two strategies. Finally, we combine a general flexible parametric structure with a mixture of two strategies. We make use of the predictive density combination scheme in Section 2 in order to obtain the time-varying weights of selected model and investment strategy combinations.

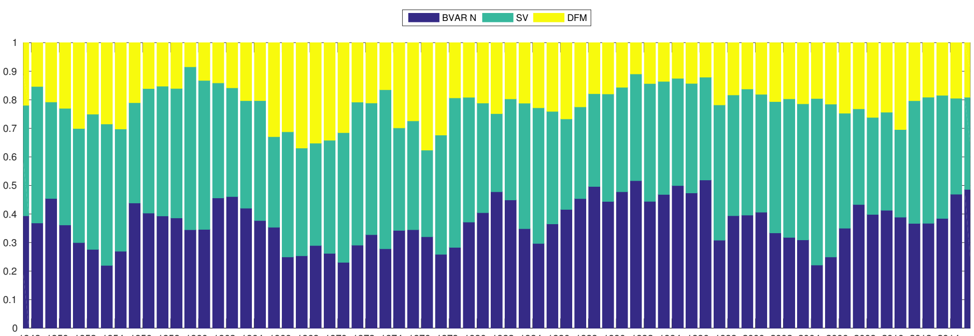
Combining basic model structures and investment strategies

In order to identify the importance of different model structures, we consider a mixture of three basic models: a VAR-N model which takes the stationary time series pattern of the series into account, a SV model which takes the time varying volatility into account, and a DFM model which captures cross-sectional correlation through common factors for series. These three models constitute the elements of the most general model we consider, namely the FAVAR-SV model. Assessing the time-varying weights of these elements presents evidence on the effect of the time variation in the stylized facts on model specification. Similar to the earlier section, each model is combined with two investment strategies, leading to 12 sets of models and investment strategy combinations. For the DFM model,

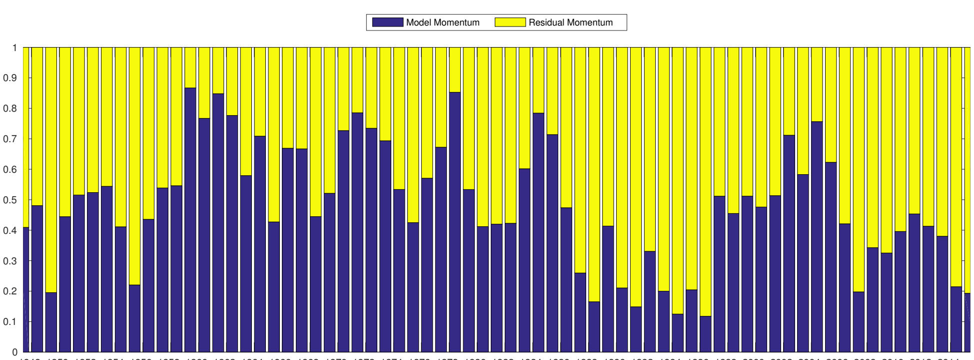
Figure 3: *Posterior weights from the six-component mixture of three basic models (VAR-N, SV and DFM(4,2)) and two strategies (model momentum and residual momentum). The figure presents posterior means of all component weights in Figure 3(a), cumulative weights for each model in Figure 3(b) and cumulative weights for the two investment strategies in Figure 3(c).*



(a) All Weights



(b) Cumulative Model Weights



(c) Cumulative Strategy Weights

we consider the general case of 4 factors and 2 lags in the factor equation.
For clarity, we present the specific density combination scheme from (12):

$$f(r_t|\mathrm{I}_{\mathbf{K}}) = \sum_{m=1}^{3} \sum_{s=1}^{2} w_{m,s,t} \int_{\mathbb{R}} f(r_t|\tilde{r}_{m,s,t}, I_k) f(\tilde{r}_{m,s,t}|I_k) d\tilde{r}_{m,s,t}, \tag{18}$$

where $m = 1, 2, 3$ is the model indicator corresponding to the models VAR-N, SV, DFM(4,2) and $s = 1, 2$ is the strategy indicator for M.M. and R.M.

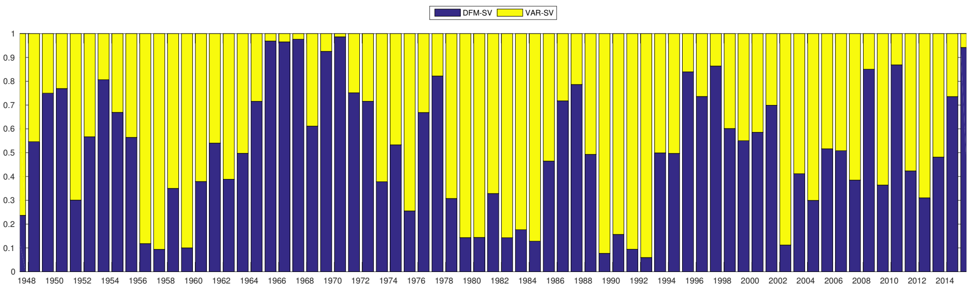
Posterior weights of each model and investment strategy are shown in Figure 3. Figure 3(a) shows the posterior mean weights for 6 mixture components arising from three models and two investment strategies. Model and strategy weights show clear time variation suggesting that autocorrelation, cross-correlation and time varying volatility patterns are important to incorporate, but the importance of these stylized facts for predictive return densities changes over time.

Figure 3(b) presents the decomposition of time-varying weights with respect to the model classes. The weights for all models change over time. In particular, VAR and DFM models have pronounced time-variation in the weights. Hence volatility properties and data reduction through common factors are more important in specific time periods. The autoregressive component captured by the VAR-N model is particularly important around 1990s and recently after 2008. Figure 3(c) shows that the investment strategies also have changing weights over time. These changes are more pronounced than those for the models. R.M. strategies are more important particularly around 1990s and at the beginning of the recent financial crisis around 2008.

Summarizing features of returns and risk are presented in Table 6. The top panel summarizes the return and risk measures obtained from model and strategy combinations, specifically from combining three models and two strategy options. The bottom panel in Table 6 presents the return and risk features when investment strategies are combined in each model, VAR-N, SV and DFM-SV. Returns from these combination of strategies are calculated using the normalized weights for two momentum strategies in each model.

Together with the dynamic weight patterns, Table 6 leads to the following conclusions in terms of the *return and risk features*. The top panel of Table 6 shows that a mixture of two strategies combined with a mixture of three basic models of the FAVAR class leads to improved risk features. Volatility of returns together with the largest loss of the combined model is typically lower then those of the individual model and strategy combinations in the top two panels of Table 5. Such improvement in the risk features stems from both components of the mixture, namely model as well as strategy mixture: Strategy mixtures in the bottom panel of Table 6 lead to better risk features compared to the individual components in Table 5, while model and strategy mixtures in the top panel of Table 6 lead in most cases to lower variances compared to the strategy mixtures in the bottom panel

Figure 4: *Posterior weights from two flexible models (DFM-SV and VAR-SV) and a mixture of investment strategies. The figure presents posterior means of model weights. DFM-SV in this figure denotes the 'combination' of DFM-SV models with $K = 1, \ldots, 4$ factors, $L = 1, 2$ lags and the two strategies. VAR-SV in this figure denotes the 'combination' of VAR-SV model and the two strategies.*



of Table 6.

Despite the improvements in risk features, model and strategy mixtures do not give substantially better mean returns and Sharpe ratios compared to the individual models and strategies in Table 5 or the mixture of strategies in the bottom panel of Table 6. One reason is that all model and strategy mixtures lead to relatively wide posterior densities for returns, as indicated by the 90% credible intervals. This relatively large uncertainty in mean returns is also reflected in the relatively wide credible intervals for the Sharpe ratios. Similar to mean returns, neither mixture of models nor mixture of strategies lead to substantially better Sharpe ratios. This is apparently due to the weight of DFM-N(4,2) model. As shown in Figure 3 the weight of the so-called "bad" model DFM-N(4,2) with M.M. remains substantial in the combination schemes of strategies. It is somewhat reduced in the mixture of models and strategies as shown in the top panel.

Thus, a combination of a mixture of two strategies with a mixture of three models leads to better risk features than for the cases where a mixture of strategies is combined with individual models and separate strategies are combined with separate models. This suggests that complete densities of the combined strategies contain relevant information. Another conclusion is that a "bad" component in a mixture of models should be a candidate for removal.

Combining a mixture of two flexible models and a mixture of investment strategies

Given the diversity in results of the mixture of the three basic model components it may be a good strategy to exclude a "bad" component and explore a smaller mixture of more flexible models. Here we explore a mixture of a VAR-SV and a DFM-SV, where for

Table 7: *Returns and risk from a mixture of two flexible models and two investment strategies. The top panel shows the return and risk from two model combinations (VAR-SV, DFM-SV(1-4, 1-2)) combined with two investment strategies (M.M. R.M.). The bottom panel reports these results for the mixture of two investment strategies combined with each of the two models separately. Standard momentum strategy has mean 0.09, volatility 5.7, Sharpe ratio 0.02 and largest loss -26.2. Bold values indicate an 'equal or better' value compared to standard momentum. 90% credible intervals are reported in parentheses.*

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

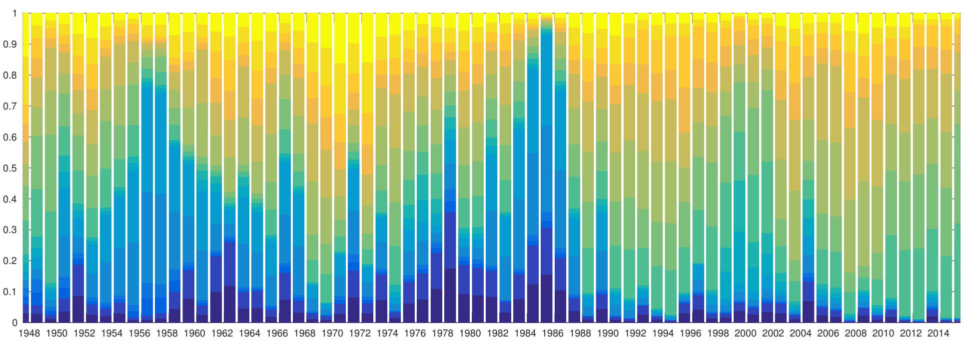
the latter model we consider the case where one uses the density combination scheme to optimize over the number of factors from 1 to 4 and number of lags from 1 to 2. In Table 7 and Figure 4, we present results of these combinations of mixtures of model and strategies in order to assess both the effect of reducing the number of components as well as the time-varying degree of possible dimension reduction, i.e. the time-varying pattern in the number of factors that explain the 10 series.

Figure 4 shows a similar pattern as Figure 3 in the time-varying behavior of the model weights. The summarizing features of return and risk, shown in Table 7, indicate that a combination of a mixture of models and strategies leads to better results than presented in the top panel of Table 5. The results for the mixture of strategies per model indicate that for VAR-SV the return features are good but not the risk features while for the case of the model DFM-SV(1:4, 1:2) the opposite holds. If an investor is interested in the joint good behavior of return and risk than averaging is beneficial.

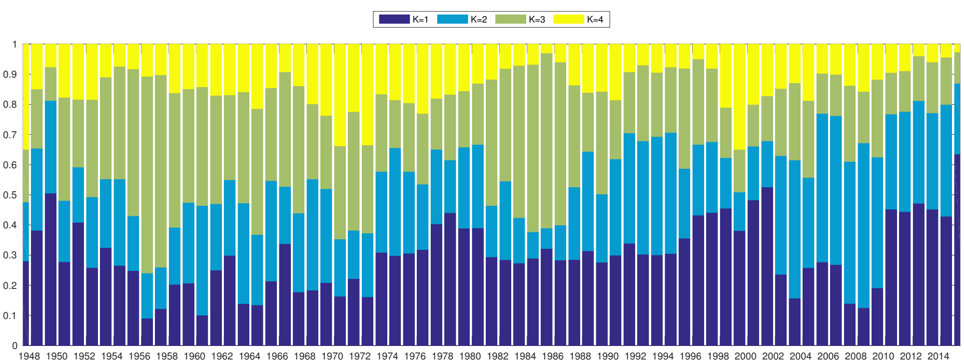
Combining a very flexible parametric model and a mixture of investment strategies

We finally explore the return and risk features of a flexible parametric structure, specified as a FAVAR-SV models averaged over 4 factors and 2 lags, combined with two investment strategies, leading to 16 predictive densities. The specific combination scheme in this case,

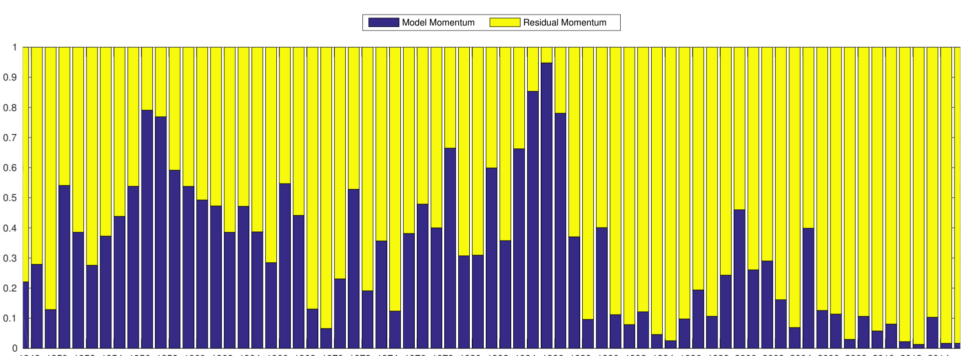
Figure 5: *Posterior weights from DECO combination of 16 FAVAR-SV model and strategy combinations: all model weights, cumulative weights of each strategy and cumulative weights of number of factors in the model. The figure presents posterior means of all component weights in Figure 5(a), cumulative weights for different number of factors in Figure 5(b) and cumulative weights for the two investment strategies in Figure 5(c).*



(a) All Weights



(b) Cumulative Weights for Number of Factors



(c) Cumulative Strategy Weights

Table 8: *Returns and risk from a combination of a very flexible parametric model and two investment strategies. The top panel shows the returns and risk from the very flexible model (FAVAR-SV(1-4, 1-2)) and a mixture of two investment strategies (M.M. R.M.). The bottom panel reports these results for the mixture of two investment strategies combined with each model separately. Standard momentum strategy has mean 0.09, volatility 5.7, Sharpe ratio 0.02 and largest loss -26.2. Bold values indicate an 'equal or better' value compared to standard momentum. 90% credible intervals are reported in parentheses.*

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

in relation to the general equation (12), is as follows:

$$f(r_{kt}|\mathrm{I}_{\mathbf{K}}) = \sum_{m=1}^{4}\sum_{l=1}^{2}\sum_{s=1}^{2} w_{m,l,s,t} \int_{\mathbb{R}} f(r_t|\tilde{r}_{m,l,s,t}, I_k) f(\tilde{r}_{m,l,s,t}|I_k) d\tilde{r}_{m,l,s,t}, \tag{19}$$

where all combined models are member of that FAVAR-SV class and $m = 1, \ldots, 4$ is the number of factors, $l = 1, 2$ is the number of AR components in each factor, $s = 1, 2$ is the strategy indicator for M.M. and R.M. and the rest of the parameters are defined as in equation (9).

Figure 5(a) shows that the weights of different FAVAR-SV models and combined strategies change substantially over time. No single model-strategy pair from the considered specifications seems to be suitable for these data. The cumulative weights per number of factors, calculated from the posterior means, are presented in Figure 5(b). These weights also change over time indicating time variation in the number of factors and hence in the amount of data reduction. Particularly in the more recent periods the models and strategy combinations with a single factor have higher weights than in earlier periods. This finding is in line with the relatively low canonical correlations between returns at the end of the sample as compared to the beginning of the sample, as shown in Figure 1(b). Time variation in the investment strategies is shown in Figure 5(c). Weights of the two investment strategies change drastically over time. Interestingly, M.M. appears to be important until the 1990s, while R.M. gains importantance in later periods.

We next report the risk and return properties from the combination of the FAVAR-SV models and the two investment strategies in Table 8. The top panel shows the returns and risk from the very flexible model (FAVAR-SV(1:4, 1:2)) and a mixture of two investment strategies (M.M. R.M.). The bottom panel reports these results for the mixture of two investment strategies for each model separately. Results for each model and each strategy were presented in Table 5. By comparing the results from Table 5 with the bottom panel of Table 8, it is seen that the mixture of strategies leads to a positive return and lower risk than for the individual strategies. The top panel shows that the using our density combination scheme to optimally average over factors and lags leads to further improvement in return features but not risk.

Broad empirical conclusions

Conditional upon our information set that consists of US industry portfolios between 1926M7 and 2015M6, a set of dynamic models and, further, two equity momentum strategies, the results of this section lead to the following broad 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.

6 Conclusions

We have presented a dynamic asset-allocation approach which combines alternative models and momentum strategies and in which the portfolios and the portfolio strategies are updated every decision period. To capture the short-run and the long-run portfolio dynamics we have specified a class of flexible dynamic models with different number of latent factors, a vector autoregression and a stochastic volatility component. Next, we have combined the models from the introduced class with a set of portfolio strategies. To this end, we have extended the density combination scheme of Casarin et al. (2016) to obtain mixtures of model structures and momentum strategies at each time period. Finally, to perform inference in the resulting nonlinear non-Gaussian state space model, we have devised a novel and efficient nonlinear non-Gaussian filter, the M-Filter. It is, based on the MitISEM algorithm of Hoogerheide et al. (2012) and leads to substantial accuracy and speed gains.

Our empirical results on the US industry portfolios from 1926-2015 indicate that time-varying combinations of the flexible models from the FAVAR-SV class and of the two strategies lead to better return and risk features than both very simple and very complex models. More specifically, flexible model combinations help to improve the most important return features, e.g. mean returns and Sharpe ratios, while combinations of two strategies help to reduce the risk features, such as volatility and the largest loss. The latter result indicates that complete densities provide useful information for risk. These results are obtained given typical data features like stationary return patterns, strong time-varying cross correlations and stochastic volatilities. The importance of these data features changes substantially over time for the data studied.

There are several opportunities to extend this line of research. Using a larger data set of industrial portfolios, more involved strategies, improved learning and a sub-period analysis to explore the results in more detail are all topics for further research.

Appendix A

In this section we describe different model structures used in Section 2 resulting from the general formulation (1).

Linear and Gaussian Dynamic Factor Model (DFM)

The linear Gaussian DFM is a special case of equation (1) with $\boldsymbol{\beta} = 0$ and a diagonal Σ matrix:

$$
\begin{aligned}
\mathbf{y}_t &= \Lambda \mathbf{f}_t + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t &\sim \mathrm{N}(0, \Sigma), \\
\mathbf{f}_t &= \phi_1 \mathbf{f}_{t-1} + \ldots + \phi_L \mathbf{f}_{t-L} + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim \mathrm{N}(0, \mathrm{Q}),
\end{aligned} \tag{20}
$$

To estimate this model we assume the following priors.

1) The diagonal elements of Σ have independent Inverse Gamma (IG) priors:

$$\underline{\sigma}^2_{\varepsilon,ii} \sim \mathrm{IG}\left(\frac{v_i}{2}, \frac{s_i}{2}\right),$$

where we set $v_i = 2$ and $s_i = 5$ for $i = 1, \ldots, N$.

2) The loading parameters have normal priors, $\underline{\Lambda} \sim \mathrm{N}(\underline{\boldsymbol{\mu}}, \underline{\boldsymbol{C}})$ where $\underline{\boldsymbol{\mu}} = 0$ and $\underline{\boldsymbol{C}} = \mathrm{I}$.

3) The prior for the autoregressive parameters $\Phi = [\phi_1, \ldots, \phi_L]$ and latent errors variance Q are diffuse conjugate Normal-Wishart:

$$\underline{\boldsymbol{\Phi}}|\mathrm{Q} \sim \mathrm{N}(0, \mathrm{Q} \otimes \Omega_0), \quad \underline{\mathrm{Q}} \sim \mathrm{iW}(\mathrm{Q}_0, N + K + 2),$$

where $\boldsymbol{\Phi} = \mathrm{vec}(\Phi)$ consists of the elements of Φ stacked in a column vector of length $L \times K^2$, where L is the number of lags of the latent factor and K is the number of factors. As in Bernanke et al. (2005) we set the prior to express the beliefs that parameters on longer lags are more likely to be zero, in the spirit of the Minnesota prior. The diagonal elements of Q_0 are set to the residual variances of the corresponding univariate autoregressions, $\hat{\sigma}^2_{\eta,kk}$ for $k = 1, \ldots, K$. The diagonal elements of Ω_0 are set on k lagged j'th variable in i'th equation equals $\sigma_i^2/k\sigma_j^2$.

Defining $\Lambda_i = (\lambda_{i,1}, \ldots, \lambda_{i,k})$ for $i = 1, \ldots, N$ the Gibbs sampling steps are as follows.

1) The full conditional posterior for the elements of Σ reduces to a set of N independent inverse-gamma distributions with:

$$\overline{\sigma}^2_{\varepsilon,ii} \sim \mathrm{IG}\left(\frac{v_i + T}{2}, \frac{v_i s_i^2 + d_i}{2}\right),$$

where $d_i = \sum_{t=1}^{T}(y_{it} - \Lambda_i f_{it})(y_{it} - \Lambda_i f_{it})', \quad i = 1, \ldots, \mathrm{N}.$

2) The draws of the loading parameters which satisfy the related restrictions are generated as follows.

   a) For $i = 1, \ldots, K$, draw $\overline{\Lambda}_i \sim \mathrm{N}(\overline{m}_i, \overline{\mathrm{C}}_i)\mathbf{I}(\lambda_{ii} > 0)$, where $\overline{m}_i = \overline{\mathrm{C}}_i(\underline{\mathrm{C}}_i^{-1}\underline{\mu}_i + \sigma_{\varepsilon,ii}^{-2} f_i' y_i)$ and $\overline{\mathrm{C}}_i^{-1} = \underline{\mathrm{C}}_i^{-1} + \sigma_{\varepsilon,ii}^{-2} f_i' f_i$

   b) For $i = K+1, \ldots, N$ draw $\overline{\Lambda}_i \sim \mathrm{N}(\overline{m}_i, \overline{\mathrm{C}}_i)$ where $\overline{m}_i = \overline{\mathrm{C}}_i(\underline{\mathrm{C}}_i^{-1}\underline{\mu}_i + \sigma_{\varepsilon,ii}^{-2} f_i' y_i)$ and $\overline{\mathrm{C}}_i^{-1} = \underline{\mathrm{C}}_i^{-1} + \sigma_{\varepsilon,ii}^{-2} f_i' f_i$

3) The posterior of $\mathbf{\Phi}$ and Q follows from the standard VAR form that we adopt, which can be estimated equation by equation to yield the following simulation scheme.

   a) Draw $\overline{\mathrm{Q}}$ from $\mathrm{iW}(\hat{\mathrm{Q}}, T+K+N+2)$, where $\hat{\mathrm{Q}} = \underline{\mathrm{Q}} + \hat{\Gamma}'\hat{\Gamma} + \hat{\Phi}'[\Omega_0 + (\hat{\mathrm{F}}_t'\hat{\mathrm{F}}_t)^{-1}]^{-1}\hat{\Phi}$ and $\hat{\Gamma}$ is the matrix of OLS residuals.

   b) Draw $\overline{\mathbf{\Phi}}$ from the conditional normal distribution of the form:

$$\overline{\mathbf{\Phi}} \sim \mathrm{N}(\mathrm{vec}(\tilde{\Phi}), \mathrm{Q} \otimes \tilde{\Omega}), \tag{21}$$

   where $\tilde{\Phi} = \tilde{\Omega}(\hat{\mathbf{f}}_{t-1}'\hat{\mathbf{f}}_{t-1})\hat{\Phi}$ and $\tilde{\Omega} = (\Omega_0^{-1} + \hat{\mathbf{f}}_{t-1}'\hat{\mathbf{f}}_{t-1})^{-1}$.

5) Draws the latent states $\mathbf{f}_t$ using the FF-BS algorithm as described in Carter and Kohn (1994).

6) Go to step 1.

Linear Dynamic Factor Model with Stochastic Volatility (DFM-SV)

We obtain the DFM-SV by setting $\boldsymbol{\beta} = 0$ in equation (1):

$$\begin{array}{ll} \mathbf{y}_t = \Lambda \mathbf{f}_t + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t \sim \mathrm{N}(0, \Sigma_t), \\ \mathbf{f}_t = \phi_1 \mathbf{f}_{t-1} + \ldots + \phi_L \mathbf{f}_{t-L} + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t \sim \mathrm{N}(0, \mathrm{Q}), \end{array} \tag{22}$$

and specifying a time-varying variance-covariance matrix:

$$\Sigma_t = \begin{pmatrix} \sigma_{11,t}^2 & 0 & \ldots & 0 \\ 0 & \sigma_{22,t}^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_{NN,t}^2 \end{pmatrix}, i = 1, \ldots, \mathrm{N}. \tag{23}$$

We assume that the log volatilities $h_{it} = \log(\sigma^2_{ii,t})$ follows a stationary and mean reverting process

$$h_{it} = \mu_i + \psi_i h_{it-1} + \zeta_t, \quad \zeta_t \sim \mathrm{N}(0, \gamma_{ii}), \quad \psi_i \in (-1, 1), \quad \mu_i \in \mathbb{R}.$$

Starting from equation (22) and rearranging, we get $\boldsymbol{\varepsilon}_t = \mathbf{y}_t - \Lambda \mathbf{f}_t = \mathbf{y}^*_t$. Taking the squares plus an offset constant we obtain

$$\begin{aligned}
\mathbf{y}^{**}_t &= \log[(\mathbf{y}^*_t)^2 + \bar{c}], \\
\mathbf{y}^{**}_t &= 2\mathbf{h}_t + \boldsymbol{e}_t, \\
\mathbf{h}_t &= \boldsymbol{\mu} + \boldsymbol{\psi} \mathbf{h}_{t-1} + \xi_t, \quad \xi_t \sim \mathrm{N}(0, \Gamma),
\end{aligned} \tag{24}$$

where $\boldsymbol{e}_t = \log(\boldsymbol{\varepsilon}_t)$ follows the $\chi^2(1)$ distribution. Therefore, the standard Kalman filter and smoother cannot be adopted, cf. Carter and Kohn (1994). To solve this problem Kim et al. (1998) employ a data augmentation approach and introduce a new state variable $\mathbf{s}_T = \{\mathbf{s}_1, \dots, \mathbf{s}_t\}$, so that the linear, non-Gaussian state space model (24) can be rewritten as conditionally linear Gaussian. Then, the distribution of $\boldsymbol{e}_t$ can be approximated as

$$\boldsymbol{e}_t \approx \sum_{j=1}^{7} q_j \mathrm{N}(\tau_j - 1.2704, \nu_j^2),$$

where τ_j, ν_j^2 and q_j for $j = 1, \dots, 7$ are constant specified in Kim et al. (1998). Conditionally on the state $s_{t+1} = j$, the errors $\boldsymbol{e}_t$ can be sampled as

$$\begin{aligned}
\boldsymbol{e}_t | \mathbf{s}_{t+1} = j &\sim \mathrm{N}(\tau_j - 1.2704, \nu_j^2), \\
\Pr(\mathbf{s}_{t+1} = j) &= q_j.
\end{aligned}$$

The draws from the sequence of states $\mathbf{s}_t$ can be obtained by using:

$$\Pr(\mathbf{s}_t = j | \mathbf{y}^{**}_t, \mathbf{h}_t) \propto q_j f_{\mathrm{N}}(\mathbf{y}^{**}_t | 2\mathbf{h}_t + \tau_j - 1.2704, \nu_j^2), \tag{25}$$

where $f_{\mathrm{N}}(\cdot)$ denotes the kernel of a normal density and $j = 1, \dots, 7$, $t = 1, \dots, T$. Conditional on $\mathbf{s}_{1:T}$ the model is linear Gaussian and the algorithm of Carter and Kohn (1994) can be used.

The priors remains as described before, with the only difference related to the SV parameters, $\boldsymbol{\mu}$, $\boldsymbol{\psi}$ and variance of the errors Γ. For the the two former we specify

$$\begin{aligned}
\begin{pmatrix} \mu_i \\ \psi_i \end{pmatrix} &\sim \mathrm{N}\left[\begin{pmatrix} \underline{\mathrm{m}}_{\mu_i} \\ \underline{\mathrm{m}}_{\psi_i} \end{pmatrix}, \begin{pmatrix} \underline{\mathrm{V}}_{\mu_i} & 0 \\ 0 & \underline{\mathrm{V}}_{\psi_i} \end{pmatrix} \right], \\
\psi_i &\in (-1, 1),
\end{aligned}$$

while for γ_{ii}^{-2} we put

$$\gamma_{ii}^{-2} \sim \mathrm{G}(1/\underline{\mathrm{k}}_{\gamma}, 1).$$

For the hyperparameters we follow Pettenuzzo and Ravazzolo (2016) and set $\underline{\mathrm{k}}_{\gamma} = 0.01$, $\underline{\mathrm{m}}_{\mu_i} = 0$, $\underline{\mathrm{m}}_{\psi_i} = 0.95$, $\underline{\mathrm{V}}_{\mu_i} = 10$ and $\underline{\mathrm{V}}_{\psi_i} = 1.0e^{-06}$. These values imply a strong autocorrelation structure for h_{it}, which is typical for financial time series.

For this model, the Gibbs sampling steps are as follows.

1) Initialize $\mathbf{f}_t^{(0)}$, $\mathbf{h}_t^{(0)}$, $\Lambda_t^{(0)}, \Sigma^{(0)}, \mathrm{Q}^{(0)}$.
2) Draw latent factors $\mathbf{f}_t$ from $p(\mathbf{f}_t|, \Lambda, \mathrm{Q}, \Sigma_t, \mathbf{h}_t)$ using the FF-BS algorithm described in Carter and Kohn (1994).
3) Conditionally on $\mathbf{h}_t$ and Λ, draw the indicator variable $\mathbf{s}_t$ for the mixture according to (25).
4) Draw a sequence of stochastic volatilities $\mathbf{h}_t$, $t = 1, \dots, T$ from $p(\mathbf{h}_t|\Lambda, \mathbf{f}_t, \mathbf{s}_t, \boldsymbol{\mu}, \boldsymbol{\psi})$ from the conditional linear and Gaussian system using the method of Carter and Kohn (1994).
5) Draw the stochastic volatility variances γ_{ii}^2 from $p(\gamma_{ii}^2|h_{it}, \mu_i, \psi_i)$ from the following posterior:

$$\overline{\gamma}_{ii}^{-2} \sim \mathrm{G}\left(\left[\frac{\underline{\mathrm{k}}_{\gamma} + \sum_{t=1}^{T-1}(h_{it+1} - \mu_i - \psi_i h_{it})^2}{t}\right]^{-1}, T\right).$$

6) Draw the SV parameters jointly:

$$\left(\begin{array}{c} \overline{\mu}_i \\ \overline{\psi}_i \end{array}\right) \sim \mathrm{N}\left(\left[\begin{array}{c} \bar{m}_{\mu_i} \\ \bar{m}_{\psi_i} \end{array}, \right] \bar{V}_{(\mu_i,\psi_i)}\right) \times \psi_i \in (-1, 1),$$

where

$$\bar{V}_{(\mu_i,\psi_i)} = \left\{\left[\begin{array}{cc} \underline{\mathrm{V}}_{\mu_i}^{-1} & 0 \\ 0 & \underline{\mathrm{V}}_{\psi_i}^{-1} \end{array}\right] + \bar{\gamma}_{ii}^{-2} \sum_{t=1}^{T-1} \left[\begin{array}{c} 1 \\ h_{it} \end{array}\right] \left[1 \quad h_{it}\right]\right\}$$

and

$$\left[\begin{array}{c} \bar{m}_{\mu_i} \\ \bar{m}_{\psi_i} \end{array}\right] = \bar{V}_{(\mu_i,\psi_i)} \left\{\left[\begin{array}{cc} \underline{\mathrm{V}}_{\mu_i}^{-1} & 0 \\ 0 & \underline{\mathrm{V}}_{\psi_i}^{-1} \end{array}\right] \left[\begin{array}{c} \underline{\mathrm{m}}_{\mu_i} \\ \underline{\mathrm{m}}_{\psi_i} \end{array}\right] + \bar{\gamma}_{ii}^{-2} \sum_{t=1}^{T-1} \left[\begin{array}{c} 1 \\ h_{it} \end{array}\right] h_{it+1}\right\}.$$

7) Go to step 2.

Linear Dynamic Factor Model with Two Stochastic Volatility Components (DFM-SV2)

We obtain the DFM model with two stochastic volatilities by assuming $\boldsymbol{\beta} = 0$ in equation (1) and by defining the following time-varying covariance matrices for the idiosyncratic and latent errors:

$$\begin{aligned} \mathbf{y}_t &= \Lambda \mathbf{f}_t + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t &\sim \mathrm{N}(0, \Sigma_t), \\ \mathbf{f}_t &= \phi_1 \mathbf{f}_{t-1} + \ldots + \phi_L \mathbf{f}_{t-L} + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim \mathrm{N}(0, \mathrm{Q}_t), \end{aligned} \tag{26}$$

with the idiosyncratic errors defined as in equation (23) and latent error variances is given by:

$$\mathrm{Q}_t = \begin{pmatrix} \eta^2_{11,t} & 0 & \ldots & 0 \\ 0 & \eta^2_{22,t} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \eta^2_{KK,t} \end{pmatrix}, i = 1, \ldots, K, \tag{27}$$

where log volatilities $k_{it} = \log(\eta^2_{ii,t})$ follow a stationary and mean reverting process:

$$k_{it} = \omega_i + \beta_i k_{it-1} + \xi_{it}, \quad \xi_{it} \sim \mathrm{N}(0, \sigma^2_{\xi_i}).$$

The estimation of this model proceeds as before with an added step in the Gibbs sampler to extract the latent time-varying variance.

Factor Augmented VAR models with Stochastic Volatility Components (FAVAR-SV2)

Assuming in equation (1) $\boldsymbol{\beta} \neq 0$ and a time-varying variance-covariance matrix for the idiosyncratic and latent errors we obtain the FAVAR-SV2 model given by:

$$\begin{aligned} \mathbf{y}_t &= \beta \mathbf{x}_t + \Lambda \mathbf{f}_t + \boldsymbol{\varepsilon}_t & \boldsymbol{\varepsilon}_t &\sim \mathrm{N}(0, \Sigma_t), \\ \mathbf{f}_t &= \phi_1 \mathbf{f}_{t-1} + \ldots + \phi_L \mathbf{f}_{t-L} + \boldsymbol{\eta}_t & \boldsymbol{\eta}_t &\sim \mathrm{N}(0, \mathrm{Q}_t). \end{aligned} \tag{28}$$

The FAVAR model extends the state equation by defining $\mathbf{x}_t$ as a vector of the lagged dependent variables. This leads to a VAR form in the state equation of (28):

$$\begin{pmatrix} \mathbf{f}_t \\ \mathbf{x}_t \end{pmatrix} = \tilde{\Phi}_1 \begin{pmatrix} \mathbf{f}_{t-1} \\ \mathbf{x}_{t-1} \end{pmatrix} + \ldots + \tilde{\Phi}_L \begin{pmatrix} \mathbf{f}_{t-L} \\ \mathbf{x}_{t-L} \end{pmatrix} + \tilde{\boldsymbol{\varepsilon}}_t,$$

see also Stock and Watson (2005). Conditional on the latent states, the estimation of the VAR parameters $\boldsymbol{\beta}$ is similar to that of the univariate linear regression model, hence Bayesian inference is standard. The two proposed FAVAR models are defined by a stochastic volatility component in the idiosyncratic disturbances (FAVAR-SV) and a stochastic volatility components in the idiosyncratic and latent disturbances (FAVAR-SV2). We refer to the earlier sections of this appendix for the inference on the SV components conditionally on the remaining parameters.

Appendix B - The M-Filter

Below we present the details of the recursion for the proposed M-Filter from Section 4. In the description we treat the estimated parameter $\hat{\theta}$ as known and we omit it for the sake of notation. For a detailed discussion of the general MitISEM procedure see Hoogerheide et al. (2012).

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 the candidate density $\tilde{g}_t(\tilde{\boldsymbol{\alpha}}_t^{(j)}|\boldsymbol{\alpha}_{t-1}^{(j)})$ using the MitISEM algorithm, which can be summarize as follows:

   a) **Initialization**: Simulate draws $\tilde{\boldsymbol{\alpha}}_t^{(j)}$ from a 'naive' candidate distribution with density $g_n(\cdot)$ (e.g. a Student-t with $v = 5$ degrees of freedom). Compute the corresponding IS weights:

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

   where the target density kernel has the form $p(\mathbf{y}_t|\tilde{\boldsymbol{\alpha}}_t^{(j)}, \tilde{\mathbf{y}}_t)p(\tilde{\boldsymbol{\alpha}}_t^{(j)}|\boldsymbol{\alpha}_{t-1}^{(j)})$, and normalise them to $w_t^{(j)}$.

   b) **Adaptation**: Use the draws $\tilde{\boldsymbol{\alpha}}_t^{(j)}$ and the weights $\tilde{w}_t^{(j)}$ from the naive distribution $g_n(\cdot)$ to IS estimate the mean and covariance matrix of the target distribution. Use these estimates as the mode and the scale matrix of the Student-t adapted density $\tilde{g}_a(\cdot)$. Draw a sample $\tilde{\boldsymbol{\alpha}}_t^{(j)}$ from $\tilde{g}_a(\cdot)$ and compute the IS weights for this sample.

   c) Apply the the **IS weighted EM (ISEM) algorithm** given the sample from the adapted distribution and the corresponding IS weights. The output consists of the new candidate density with $h = 1$ component $\tilde{g}_t^{(h)}(\cdot)$ with the optimized parameters. Draw a new sample $\tilde{\boldsymbol{\alpha}}_t^{(j)}$ from this candidate, compute the corresponding IS weights. Calculate the coefficient of variation $CV^{(h)}$ of the weights

$\tilde{w}_t^{(j)}$, $j = 1, \ldots, M$.

d) **Iterate on the number of mixture components**. Given the current mixture of h components $\tilde{g}_t^{(h)}(\cdot)$ add the next component to the mixture in the following way.

  d.1) Use a chosen fraction (e.g. 0.1) of the draws $\tilde{\boldsymbol{\alpha}}_t^{(j)}$ from the current mixture corresponding to the highest IS weights to IS estimate the mean and variance. Use these parameters as the starting mode and scale parameters for the new mixture component, μ_{h+1} and Σ_{h+1}.

  d.2) Update the mixture probabilities: assign the starting value for the new component probability η_{h+1} (e.q. 0.1) and multiply the old mixture probabilities $\eta_1, \ldots, \eta_h$ by $(1 - \eta_{h+1})$. Set the number of degrees of freedom for the new component ν_h to a specified fixed value (e.g. 5).

  d.3) Given the staring parameters of the new mixture , adapt the candidate for the model parameters by performing ISEM based on the draws from the previous mixture $\tilde{g}_t^{(h)}(\cdot)$ and the corresponding weights.

  d.4) Draw $\tilde{\boldsymbol{\alpha}}_t^{(j)}$ from the new mixture $\tilde{g}_t^{(h+1)}(\cdot)$ and evaluate the corresponding importance weights $\tilde{w}_t^{(j)}$, $j = 1, \ldots, M$.

  d.5) Calculate the coefficient of variation $CV^{(h+1)}$ of the weights $\tilde{w}_t^{(j)}$, $j = 1, \ldots, M$.

e) **Assess convergence** of the candidate density's quality by inspecting the relative change between $CV^{(h)}$ and $CV^{(h+1)}$ is greater than the chosen threshold (e.g. 0.01) and return to step d) unless the algorithm has converged.

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

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

4) **Likelihood Approximation**. The approximation of the log likelihood function is given by:

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

We finally note that the MitISEM approximation can be robustified as outlined in Baştürk et al. (2016). Especially for complex model structures and large number of variables, leading to highly multi-modal target densities, it is likely that some components of the

MitISEM approximation have very low weights and are not updated efficiently. In such cases, an additional rejection step in MitISEM to avoid components with very small mixture weights is shown to improve the accuracy and the speed of the algorithm.

Appendix C- Simulation results for M-filter

In this Appendix we report some simulation results for the M-Filter, in all the examples we are interested in the estimation of the target function $h_t(\boldsymbol{\alpha}_t^{(j)}) = \boldsymbol{\alpha}_t$ that is the posterior mean of the latent state. We compare four filters, the Kalman filter (KF), the Particle Filter (PF), the Auxiliary Particle filter (APF) and the M-Filter.[4] All Monte Carlo experiments presented in this section are based on I = 100 replications with T = 100 observations each. For the PF, APF and M-Filter we use $M = 50,000$ particles.

The first model we consider is a standard local level model:

$$\begin{aligned} y_t &= \alpha_t + \varepsilon_t \qquad \varepsilon_t \sim \mathrm{N}(0, \sigma_\varepsilon^2), \\ \alpha_t &= \alpha_{t-1} + \eta_t \quad\ \ \eta_t \sim \mathrm{N}(0, \sigma_\eta^2). \end{aligned} \tag{29}$$

Since this is a linear Gaussian model it is often used as a benchmark for comparing filtering methods. In this case the KF provides the closed form for the optimal filtering distribution.

In the simulations experiments, we fix the latent state variance as $\sigma_\eta^2 = 0.1$ and we define four different levels for the state variance σ_ε^2, corresponding to four levels of the Noise to Signal Ratio (NtS): 0.1, 0.5, 1 and 2.5. We note that the exact likelihood of the model in equation (29) can be calculated using the KF, and we can compare the exact likelihood of this model with the remaining non-linear filters. This allows to assess the degree of the likelihood bias in the non-linear filters, including the proposed M-Filter.

Table 9 reports the results for the model in equation (29). KF filter is the best filter, as expected, in terms of the minimum MAE and the smallest computing time. The results of the non-linear filters, however, are in line with those of KF in terms of the MAE measures. The proposed M-Filter performs similarly to the PF and the APF but has a lower MAE in the estimate likelihood especially for smallest NtS ratio of 0.1. In all cases the computing time is lower then the PF and APF.

As the second example we consider the Stochastic Volatility (SV) model (Kim et al., 1998)

[4]We note that our proposed filter approach is related to that of Liesenfeld and Richard (2003) and Richard and Zhang (2007). An important difference is our flexible and robust procedure to choose the candidate density. We leave a comparison between the two approaches as a topic for future research.

Table 9: *Monte Carlo results for $I = 100$ replications of the linear Gaussian model* (29) *with $T = 100$, obtained with the Kalman Filter (KF), the Bootstrap Particle Filter (PF), the Auxiliary Particle Filter (APF) and MitISEM Filter (M-Filter) with 50,000 particles. The table reports the Likelihood Bias (LB) with respect to the KF, the Mean Absolute Error (MAE) defined as $MAE = 1/T \sum_{t=1}^{T} \left(1/I \sum_{i=1}^{I} |\tilde{\alpha}_{t,i} - \alpha_{t,i}| \right)$ relative to the KF, the Mean Square Error (MSE) defined as $MSE = 1/T \sum_{t=1}^{T} \left(1/I \sum_{i=1}^{I} (\tilde{\alpha}_{t,i} - \alpha_{t,i})^2 \right)$ relative to the KF. The final column reports the computing time in seconds for the four filters.*

NtS		*0.1*			*0.5*		*Time with NtS*	
Model	LB	MAE	MSE	LB	MAE	MSE	0.1	0.5
KF	0.00	1.00	1.00	0.00	1.00	1.00	0.01	0.01
PF	-48.93	1.22	1.48	-19.43	1.26	1.62	33.71	35.55
APF	-13.87	1.00	1.00	-9.56	1.01	1.02	35.54	37.67
M-Filter	-10.40	1.00	1.01	-9.52	1.01	1.02	12.83	12.81
NtS		*1*			*2.5*		*Time with NtS*	
Model	LB	MAE	MSE	LB	MAE	MSE	1	2.5
KF	0.00	1.00	1.00	0.00	1.00	1.00	0.01	0.01
PF	-37.85	1.31	1.71	-21.16	1.43	2.04	35.22	34.53
APF	-10.43	1.00	1.00	-9.05	1.00	1.00	37.29	35.72
M-Filter	-10.18	1.01	1.01	-9.39	1.00	1.01	12.67	12.13

specified as:

$$\begin{aligned} y_t &= e^{(\alpha_t/2)} \varepsilon_t & \varepsilon_t &\sim \mathrm{N}(0, \sigma_\varepsilon^2), \\ \alpha_t &= \mu + \phi \alpha_{t-1} + \eta_t & \eta_t &\sim \mathrm{N}(0, \sigma_\eta^2), \end{aligned} \tag{30}$$

where η_t and ε_t are i.i.d. and y_t is the observed series. Due to the non-linear structure of the observation equation the analytical form for filtering and predictive densities do not exist in this model.

In the simulations, we fix the autoregressive parameter ϕ to 0.90, 0.95, and 0.98, which are the typical values found in empirical studies, see among others Aguilar and West (2000). For each value of ϕ we consider four values of σ_η^2 corresponding to different coefficient of variation (CV) of the volatility $h = \bar{\sigma}^2 \exp(\alpha_t)$, which is defined as

$$\mathrm{CV} = \frac{\mathrm{Var}(h)}{\mathrm{E}(h)^2} = \exp\left(\frac{\sigma_\eta^2}{1 - \phi^2} \right) - 1.$$

We consider for the CV the values 0.1, 0.5, 1, and 2.5, where high values indicate more strength of the volatility process and low values indicate that the volatility is close to constant.

Table 10 reports the results for the SV model (30) with $\phi = 0.98$ and different values of σ_η^2. In all the cases the KF is the worst filter due to being a linear and Gaussian filter. The M-Filter performs similarly to the PF and the APF in term of the bias and estimation variability. In this model the computational speed is comparable between the three non-linear filters, namely PF, APF and M-Filter.

Table 10: *Monte Carlo results for I=100 replications of the SV model in equation* (30) *with T=100,* $\phi = 0.98$ *and* $CV = 0.1, 0.5, 1, 2.5$. *The Kalman Filter (KF), Bootstrap Particle Filter (PF), Auxiliary Particle Filter (APF) and MitISEM Filter (M-Filter) with 50,000 particles. The table reports the Likelihood Bias (LB) with respect to the KF, the Mean Absolute Error (MAE) defined as* $MAE = 1/T\sum_{t=1}^{T}\left(1/I\sum_{i=1}^{I}|\tilde{\alpha}_{t,i} - \alpha_{t,i}|\right)$ *relative to the KF, the Mean Square Error (MSE) defined as* $MSE = 1/T\sum_{t=1}^{T}\left(1/I\sum_{i=1}^{I}(\tilde{\alpha}_{t,i} - \alpha_{t,i})^2\right)$ *relative to the KF. The final column reports the computing time in seconds for the four filters for different CV.*

CV	*0.1*		*0.5*		*Time*	
Model	MAE	Var	MAE	Var	0.1	0.5
KF	1.00	1.00	1.00	1.00	0.01	0.01
PF	0.24	0.10	0.31	0.12	13.82	13.99
APF	0.25	0.10	0.31	0.13	14.58	14.66
M-Filter	0.26	0.10	0.31	0.14	14.15	12.67
CV	*1.0*		*2.5*		*Time*	
Model	MAE	Var	MAE	Var	1	2.5
KF	1.00	1.00	1.00	1.00	0.01	0.01
PF	0.32	0.12	0.29	0.11	13.98	13.88
APF	0.31	0.13	0.29	0.11	14.61	14.70
M-Filter	0.30	0.13	0.28	0.11	13.54	12.96

The last model we examine is the Dynamic Factor Model (DFM) specified as

$$\begin{aligned} \mathbf{y}_t &= \Lambda \mathbf{f}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim \mathrm{N}(0, \Sigma), \\ \mathbf{f}_t &= \Phi_1 \mathbf{f}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathrm{N}(0, \mathrm{Q}), \end{aligned} \tag{31}$$

where $\mathbf{y}_t$ is a $N \times 1$ vector of time series, a $K \times 1$ vector $\mathbf{f}_t$ contains unobservable factors with one lag where Φ_1 is a $K \times K$ matrix of autoregressive coefficients, Λ is an $N \times K$ matrix of factor loadings. Finally, $\boldsymbol{\varepsilon}_t$ is an $N \times 1$ i.i.d. vector of idiosyncratic disturbances and $\boldsymbol{\eta}_t$ is an $K \times 1$ i.i.d. vector of latent disturbances. Notice that model (31) is linear Gaussian, hence the KF is the optimal filter. As previously, we compare the performance of the non-linear filters against the KF for different number of factors.

Table 11 reports the results for the DFM model for $I = 100$ Monte Carlo replication, $N = 20$ series and $K = 2, 4, 6, 10$ factors. Obviously, the KF performs the best in this case, but the non-linear filters are not far off. The M-Filter performs better than both the PF and the APF, with substantially lower MAE and MSE. The M-Filter has also the lowest likelihood bias compared to the other non-linear and non-Gaussian filters. Computing time increases with the number of factors for all the filters, however, M-Filter requires less computing time compared to the PF and APF.

Table 11: *Monte Carlo results for I=100 replications of the DFM with T=100, $N = 20$ and $K = 2, 4, 6, 10$ latent factors for the Kalman Filter (KF), the Bootstrap Particle Filter (PF), the Auxiliary Particle Filter (APF) and the MitISEM Filter (M-Filter) with 50,000 particles. The table reports Likelihood Bias (LB) with respect to KF, the Mean Absolute Error (MAE) defined as $MAE = 1/T\sum_{t=1}^{T}\left(1/I\sum_{i=1}^{I}|\tilde{\alpha}_{t,i} - \alpha_{t,i}|\right)$ relative to the KF, Mean Squared Error (MSE) defined as $Var = 1/T\sum_{t=1}^{T}\left(1/I\sum_{i=1}^{I}(\tilde{\alpha}_{t,i} - \alpha_{t,i})^2\right)$ relative to the KF. The final column reports the computing time in seconds for the four filters and for $K = 2, 4, 6, 10$.*

Factors		*2*			*4*			*Time*
Model	LB	MAE	MSE	LB	MAE	MSE	2	4
KF	0	1	1	0	1	1	0.01	0.01
PF	-77.42	1.15	1.33	-145.49	1.15	1.32	708.79	811.73
APF	-39.98	1.03	1.05	-164.80	1.05	1.05	836.69	878.13
M-Filter	-23.23	1.01	1.02	-23.39	1.00	1.01	106.33	138.18
Factors		*6*			*10*			*Time*
Model	LB	MAE	MSE	LB	MAE	MSE	6	10
KF	0.00	1.00	1.00	0.00	1.00	1.00	0.02	0.02
PF	-193.74	1.16	1.31	-333.33	1.27	1.65	861.10	897.86
APF	-309.26	1.07	1.12	-568.18	1.08	1.18	953.72	1011.21
M-Filter	-16.97	1.03	1.03	-112.68	1.02	1.03	213.20	402.82

References

Aguilar O, West M. 2000. Bayesian Dynamic Factor Models and Portfolio Allocation. *Journal of Business & Economic Statistics* **18**: 338–357.

Asness C S, Moskowitz T J, Pedersen L H. 2013. Value and Momentum Everywhere. *The Journal of Finance* **68**: 929–985.

Bai J, Peng W P. 2015. Identification and Bayesian Estimation of Dynamic Factor Models. *Journal of Business & Economic Statistics* **33**: 221–240.

Baştürk N, Grassi S, Hoogerheide L, Opschoor A, van Dijk H K. 2016. The R-package MitISEM: Efficient and Robust Simulation Procedures for Bayesian Inference. *Journal of Statistical Software* Forthcoming.
URL `http://www.tinbergen.nl/discussionpaper/?paper=2466`

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.

Billio M, Casarin R, Ravazzolo F, van Dijk H K. 2013. Time-Varying Combinations of Predictive Densities using Nonlinear Filtering. *Journal of Econometrics* **177**: 213–232.

Blitz D, Huij J, Martens M. 2011. Residual Momentum. *Journal of Empirical Finance* **18**: 506–521.

Carter C, Kohn R. 1994. On Gibbs Sampling for State Space Models. *Biometrika* **81**: 541–553.

Casarin R, Grassi S, Ravazzolo F, van Dijk H K. 2015. Parallel Sequential Monte Carlo for Efficient Density Combination: The DeCo MATLAB Toolbox. *Journal of Statistical Software* **68**: 1–30.

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 15–084/III, Tinbergen Institute.
URL `https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2633388`

Chan L K C, Jegadeesh N, Lakonishok J. 1996. Momentum Strategies. *The Journal of Finance* **51**: 1681–1713.

Creal D. 2012. A Survey of Sequential Monte Carlo Methods for Economics and Finance. *Econometric Reviews* **31**: 245–296.

Doucet A, de Freitas N, Gordon N. 2001. *Sequential Monte Carlo Methods in Practice.* New York, USA: Springer Verlag.

Fama E F, French K R. 1992. The Cross-Section of Expected Stock Returns. *The Journal of Finance* **47**: 427–465.

Fama E F, French K R. 1993. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* **33**: 3–56.

Fama E F, French K R. 1996. Multifactor Explanations of Asset Pricing Anomalies. *The Journal of Finance* **51**: 55–84.

Fama E F, French K R. 2015. A Five-Factor Asset Pricing Model. *Journal of Financial Economics* **116**: 1–22.

Geweke J, Zhou G. 1996. Measuring the Pricing Error of the Arbitrage Pricing Theory. *Review of Financial Studies* **9**: 557–587.

Gordon N J, Salmond D J, Smith A F M. 1993. Novel Approach to Nonlinear/non-Gaussian Bayesian State Estimation. In *IEE Proceedings F (Radar and Signal Processing)*, volume 140. IET, 107–113.

Han Y. 2006. Asset Allocation with a High Dimensional Latent Factor Stochastic Volatility Model. *Review of Financial Studies* **19**: 237–271.

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.

Jegadeesh N, Titman S. 1993. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance* **48**: 65–91.

Jegadeesh N, Titman S. 2001. Profitability of Momentum Strategies: An Evaluation of Alternative Explanations. *The Journal of Finance* **56**: 699–720.

Jordan M I, Jacobs R A. 1994. Hierarchical Mixtures of Experts and the EM Algorithm. *Journal Neural Computation* **6**: 181–214.

Jordan M I, Xu L. 1995. Convergence Results for the EM Approach to Mixtures of Experts Architectures. *Neural Networks* **8**: 1409–1431.

Kim S, Shephard N, Chib S. 1998. Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *The Review of Economic Studies* **65**: 361–393.

Kunsch H R. 2005. Recursive Monte Carlo Filters: Algorithms and Theoretical Analysis. *Annals of Statistics* **33**: 1983–2021.

Liesenfeld R, Richard J F. 2003. Univariate and Multivariate Stochastic Volatility Models: Estimation and Diagnostics. *Journal of Empirical Finance* **10**: 505–531.

Liu J S. 2001. *Monte Carlo Strategies in Scientific Computing*. New York, USA: Springer Verlag.

Lopes H F, West M. 2004. Bayesian Model Assessment in Factor Analysis. *Statistica Sinica* **14**: 41–68.

Moskowitz T J, Grinblatt M. 1999. Do Industries Explain Momentum? *The Journal of Finance* **54**: 1249–1290.

Ng S, Engle R F, Rothschild M. 1992. A Multi-Dynamic-Factor Model for Stock Returns. *Journal of Econometrics* **52**: 245–266.

Peng F, Jacobs R A, Tanner M A. 1996. Bayesian Inference in Mixtures-of-Experts and Hierarchical Mixtures-of-Experts Models with an Application to Speech Recognition. *Journal of the American Statistical Association* **91**: 953–960.

Pettenuzzo D, Ravazzolo F. 2016. Optimal Portfolio Choice Under Decision-Based Model Combinations. *Journal of Applied Econometrics* Forthcoming.

Pitt M K, Shephard N. 1999. Filtering via Simulation: Auxiliary Particle Filter. *Journal of the American Statistical Association* **94**: 590–599.

Quintana J M, Chopra V K, Putnam B H. 1995. Global Asset Allocation: Stretching Returns by Shrinking Forecasts. In *Proceedings of the ASA Section on Bayesian Statistical Science*. 199–205.

Richard J F, Zhang W. 2007. Efficient High-dimensional Importance Sampling. *Journal of Econometrics* **141**: 1385–1411.

Stock J H, Watson W M. 2005. Implications of Dynamic Factor Models for VAR Analysis. Technical report, NBER Working Paper No. 11467.

Winkler R L, Barry C B. 1975. A Bayesian Model for Portfolio Selection and Revision. *The Journal of Finance* **30**: 179–192.