

Basturk, Nalan; Grassi, Stefano; Hoogerheide, Lennart; van Dijk, Herman K.

Working Paper

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

Tinbergen Institute Discussion Paper, No. 16-099/III

Provided in Cooperation with:

Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Basturk, Nalan; Grassi, Stefano; Hoogerheide, Lennart; van Dijk, Herman K. (2016) : Time-varying Combinations of Bayesian Dynamic Models and Equity Momentum Strategies, Tinbergen Institute Discussion Paper, No. 16-099/III, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:

<https://hdl.handle.net/10419/149503>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

TI 2016-099/III
Tinbergen Institute Discussion Paper



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

*Nalan Basturk*¹

*Stefano Grassi*²

Lennart Hoogerheide^{3,5}

Herman K. van Dijk^{4,5}

¹ Maastricht University, The Netherlands;

² University of Kent, United Kingdom;

³ Faculty of Economics and Business Administration, VU University Amsterdam, The Netherlands;

⁴ Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands;

⁵ Tinbergen Institute, The Netherlands.

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

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

N. Baştürk¹, S. Grassi², L. Hoogerheide^{3,4}, and H.K. van Dijk^{3,4,5}

¹Maastricht University

²University of Kent

³Tinbergen Institute

⁴VU University Amsterdam

⁵Erasmus University Rotterdam

October 31, 2016

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. For modeling, a general class of models is specified that combines a dynamic factor and a vector autoregressive model and includes stochastic volatility, denoted by FAVAR-SV. It is an extended version of a factor-augmented vector autoregressive model, [Bernanke et al. \(2005\)](#). Next, a Bayesian strategy combination is introduced in order to deal with a set of strategies. Our 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 strategic weights to be dependent between strategies as well as over time and to further allow for strategy incompleteness. Our approach results in a combination of different portfolio strategies: a model-based and a residual momentum strategy, see [Blitz et al. \(2011\)](#). The estimation of this modeling and strategy approach can be done using an extended and modified version of the forecast combination methodology of [Casarin et al. \(2016\)](#). In the approach use is made of an implied state space structure for model and strategy weights. Given the complexity of the non-linear and non-Gaussian structure a new and efficient filter is introduced based on the MitISEM approach, see [Hoogerheide et al. \(2012\)](#). Using US industry portfolios between 1926M7 and 2015M6 as data, our empirical results indicate 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 largest loss. The latter result indicates that complete densities provide useful information for risk.

1 Introduction

Traditional factor models rely on macro or firm specific factors to explain expected pay-offs 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 Bootstrap Particle Filter (PF) and the Auxiliary Particle Filter (APF).

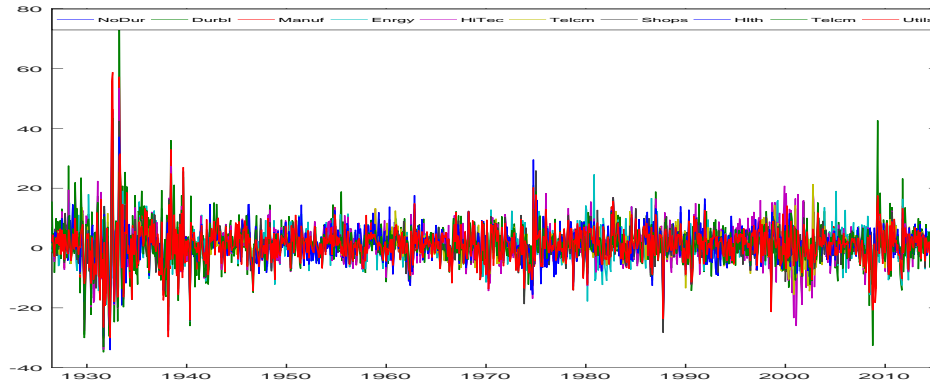
We investigate the performance of the model and portfolio strategy combination method and the M-Filter using US industry portfolios between 1926M7 and 2015M6. Our results show that time-varying combinations of flexible models in the FAVAR-SV class and two momentum strategies lead to better return and 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 MitISEM 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.

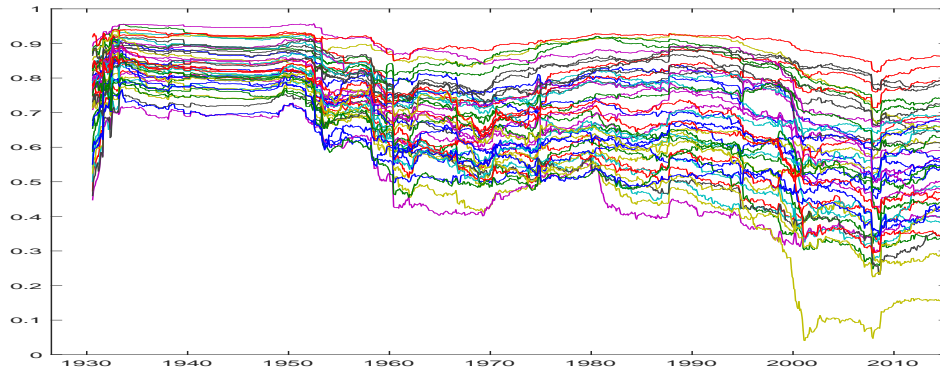
2 Stylized facts about US industrial portfolios and dynamic model structures

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

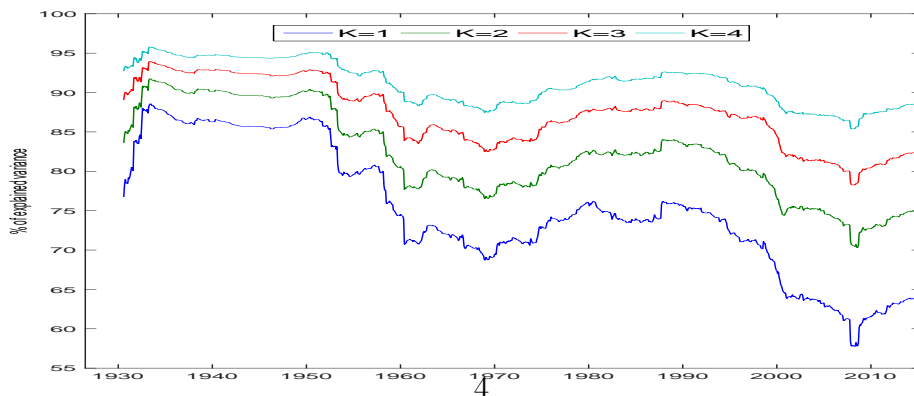
Figure 1: Monthly percentage returns, explained variation from principle components and canonical correlations 10 US industry portfolios. Principle components in Figure 1(b) and correlation calculations in Figure 1(c) are based on moving windows with 240 monthly observations. We use first 50 observations 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

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.

One may observe, at least, four stylized facts from Figures 1(a)–1(c): a stationary auto-regressive time-series pattern for all return series, a strong cross-section correlation between returns with a time-varying pattern, thirdly, a clear volatility clustering that is common to all series, and finally the total variation in the series can be captured well with one to four principle components but the explained variation, hence the number of common factors for these series, is 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:

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

In (1), the dependent variable $\mathbf{y}_t = (y_{1t}, \dots, y_{Nt})'$ 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 $t = 1, \dots, T$. 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 $j = 1, \dots, L$ 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 $N \times 1$ i.i.d. vector of idiosyncratic disturbances distributed as $\text{N}(0, \Sigma_t)$ where the time varying variance-covariance matrix may be constant as a special case and $\boldsymbol{\eta}_t$ is an $K \times 1$ i.i.d. vector of latent disturbances distributed as $\text{N}(0, \mathbf{Q}_t)$ where also the time varying variance-covariance matrix may be constant as a special case.

Different short and long-run dynamic behavior 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 $\beta = 0_{(N \times P)}$, 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 modeled. 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.

A second subclass of models has a stochastic volatility component. When this is

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

specified only in the idiosyncratic disturbances, we denote this with DFM-SV or VAR-SV. When stochastic volatility components are modeled 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 behavior 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:

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

for any $K \times K$ invertible matrix \mathbf{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 residual momentum 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 $m = 1, \dots, M$, we define a portfolio strategy \mathcal{S}_s for $s = 1, \dots, S$. 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:

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

where ι is a $1 \times P$ vector of ones, and $\mathbf{y}_{t+1:t+P} = (\mathbf{y}_{t+1} \dots \mathbf{y}_{t+P})'$ 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 $\boldsymbol{\varepsilon}_{t-P+1:t}$:

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

where $\boldsymbol{\varepsilon}_{t-P+1:t} = (\boldsymbol{\varepsilon}_{t-P+1}, \dots, \boldsymbol{\varepsilon}_t)'$ are the residuals of the model in (1), $\omega_{t,s,m} = (\omega_{t,s,m,1}, \dots, \omega_{t,s,m,N})'$ 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 $d = 1, \dots, D$, we obtain draws from the weight distribution:

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

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:

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

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:

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

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

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

3.1 Standard momentum strategy

We first summarize 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, there 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:

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

where $\bar{\mathbf{y}}_{n,t} = \sum_{p=1}^P \mathbf{y}_{n,t-p}$ and $\bar{\mathbf{y}}_t = (\bar{\mathbf{y}}_{1,t}, \dots, \bar{\mathbf{y}}_{N,t})'$ is the set of cumulative returns during the momentum period, $\mathbf{y}_t = (\mathbf{y}_{1,t}, \dots, \mathbf{y}_{N,t})'$ 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 Model based momentum and residual momentum strategies

As an alternative to the standard momentum strategy, we propose a model-based momentum strategy where the past performances of returns are evaluated according to the fitted distributions of \mathbf{y}_t , where posterior draws from the fitted distribution are obtained using MCMC. The momentum strategy in this case is similar to (7),

with the following distinction: $\bar{\mathbf{y}}_{n,t} = \sum_{p=1}^P \mathbf{y}_{n,t-p}$ and $\bar{\mathbf{y}}_t = (\bar{\mathbf{y}}_{1,t}, \dots, \bar{\mathbf{y}}_{N,t})'$ in (7) are calculated as the posterior means of the fitted return distributions. To our knowledge, such a momentum strategy is not considered in the literature, but it serves as a natural extension of the standard momentum strategy when alternative underlying models for returns are taken into account.

In addition, we consider a model-based residual momentum strategy, where the underlying model is one of the factor models in Section 2. These model and strategy combinations can be seen as an extension of [Blitz et al. \(2011\)](#).

The residual momentum strategy proposed in [Blitz et al. \(2011\)](#) sorts the returns based on past P residuals from the Fama-French factor model. The stocks with unexpectedly high past performance (corresponding to high residuals) are given a positive weight and the stocks with unexpectedly low past performance (corresponding to low residuals) are given a negative weight.

In order to make a link with our models, we explain this strategy for all models of the form (1). Given model \mathcal{M}_m , the weights are computed as follows:

$$\omega_{t,s,m,n} = \begin{cases} 1 & \text{if } \bar{\varepsilon}_{n,t}^{(m)} \geq \text{quant}(\bar{\varepsilon}_t^{(m)}, 0.90) \\ -1 & \text{if } \bar{\varepsilon}_{n,t}^{(m)} \leq \text{quant}(\bar{\varepsilon}_t^{(m)}, 0.10) \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

where $\bar{\varepsilon}_{n,t}^{(m)} = \frac{1}{P} \sum_{k=0}^{P-1} \varepsilon_{n,t-k}^{(m)}$ is the average of the estimated residual for the last P periods. In addition, $\bar{\varepsilon}_t^{(m)} = (\bar{\varepsilon}_{1,t}^{(m)}, \dots, \bar{\varepsilon}_{N,t}^{(m)})'$, $\varepsilon_t^{(m)} = (\varepsilon_{1,t}^{(m)}, \dots, \varepsilon_{N,t}^{(m)})'$ and $\text{quant}(\mathbf{x}, p)$ denotes the $100 \times p$ percent quantile of the elements of vector \mathbf{x} .

4 Density combinations and the MitISEM 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 a new filter, labelled the MitISEM Filter (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 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 predic-

tors 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 approach. The basic idea for one economic variable is as follows. The *conditional predictive probability* of an economic variable of interest y_t , given a set of K predicted variables denoted by $\tilde{\mathbf{y}}_t = (\tilde{y}_{1t}, \dots, \tilde{y}_{Kt})'$, is specified as a *discrete mixture of conditional predictive probabilities* of y_t given \tilde{y}_{kt} , $k = 1, \dots, K$ coming from K individual models or experts with weights, denoted by w_{kt} , $k = 1, \dots, K$. In terms of densities one writes:

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

where I_k with $k = 1, \dots, K$ is the information set of model k and $I_{\mathbf{K}}$ is the joint information set for all models. As a next step one 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:

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

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

$$w_{kt} = \exp\{x_{kt}\} / \sum_{k=1}^K \exp\{x_{kt}\}$$

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 be 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 $\mathbf{\Gamma}_t = \text{vec}(\mathbf{W}_t)$ 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:

$$\begin{aligned} \mathbf{y}_t &\sim p(\mathbf{y}_t | \boldsymbol{\alpha}_t, \tilde{\mathbf{y}}_t) && \text{(measurement density),} \\ \boldsymbol{\alpha}_t &\sim p(\boldsymbol{\alpha}_t | \boldsymbol{\alpha}_{t-1}, \mathbf{y}_{1:t-1}, \tilde{\mathbf{y}}_{1:t-1}) && \text{(transition density),} \\ \boldsymbol{\alpha}_0 &\sim p(\boldsymbol{\alpha}_0) && \text{(initial density).} \end{aligned} \tag{10}$$

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

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

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.

4.2 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:

$$f(r_t|I_{\mathbf{K}}) = \sum_{k=1}^K w_{kt} \int_{\mathbb{R}} f(r_t|\tilde{r}_{kt}, I_k) f(\tilde{r}_{kt}|I_k) d\tilde{r}_{kt}, \quad (11)$$

where $k = 1, \dots, K$ 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 (11) 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 notation difference between the combination schemes in (9) and the proposed one in (11), but follow the notation in (9) in the remainder of this section for a direct comparison of the proposed filtering method with the one in [Casarin et al. \(2016\)](#).

4.3 The MitISEM 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 generic state space model of equation (10) is given by the following system:

$$\begin{aligned} \mathbf{y}_t &= m_t(\boldsymbol{\alpha}_t, \boldsymbol{\varepsilon}_t), \\ \boldsymbol{\alpha}_t &= h_t(\boldsymbol{\alpha}_{t-1}, \boldsymbol{\eta}_t), \end{aligned} \quad (12)$$

where \mathbf{y}_t are the observations, $\boldsymbol{\alpha}_t$ are the state variables and $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_t$ are mutually independent errors. The state variables $\boldsymbol{\alpha}_T = \{\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_T\}$ are generally unobserved and they have to be estimated using the observed data, $\mathbf{y}_T = \{\mathbf{y}_1, \dots, \mathbf{y}_T\}$, conditional on a set of estimated parameters $\hat{\theta}$.

The object of interest is the joint 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.

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

- 1) **Initialization.** Draw the initial particles from the distribution $\alpha_0^{(j)} \sim p(\alpha_0)$ and set the weights $W_0^{(j)} = 1$ for $j = 1, \dots, M$.
- 2) **Recursion.** For $t = 1, \dots, T$
 - a.) **Forecasting α_t .** Draw $\tilde{\alpha}_t^{(j)}$ from the density $g_t(\tilde{\alpha}_t^{(j)}|\alpha_{t-1}^{(j)}, \hat{\theta})$ and define the importance weights as:

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

- b.) **Forecasting \mathbf{y}_t .** Define the incremental weights:

$$\tilde{\omega}_t^{(j)} = p(\mathbf{y}_t|\tilde{\alpha}_t^{(j)}, \hat{\theta})\omega_t^{(j)}.$$

- 3) **Updating.** Define the normalized weights

$$\tilde{W}_t^{(j)} = \frac{\tilde{\omega}_t^{(j)} W_{t-1}^{(j)}}{\frac{1}{M} \sum_{j=1}^M \tilde{\omega}_t^{(j)} W_{t-1}^{(j)}}.$$

An approximation of $E[h_t(\tilde{\alpha}_t)|\mathbf{y}_{1:t}, \theta]$ is given by:

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

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

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

The PF requires the specification of the proposal density $g_t(\tilde{\alpha}_t^{(j)}|\alpha_{t-1}^{(j)}, \hat{\theta})$, there is a large literature regarding the choice of this proposal see among others [Doucet et al. \(2001\)](#), [Liu \(2001\)](#), [Kunsch \(2005\)](#) and [Creal \(2012\)](#). The PF requires also the resampling step that can introduce additional Monte Carlo variation into the algorithm and slow down the filtering procedure.

Omitting the estimated parameters $\hat{\theta}$ to simplify notation, the crucial quantity in

PF recursion is given by:

$$p(\mathbf{y}_t|\tilde{\boldsymbol{\alpha}}_t^{(j)})\omega_t^{(j)}(\tilde{\boldsymbol{\alpha}}_t^{(j)}, \boldsymbol{\alpha}_{t-1}^{(j)}) = \frac{p(\mathbf{y}_t|\tilde{\boldsymbol{\alpha}}_t^{(j)})p(\tilde{\boldsymbol{\alpha}}_t^{(j)}|\boldsymbol{\alpha}_{t-1}^{(j)})}{g_t(\tilde{\boldsymbol{\alpha}}_t^{(j)}|\boldsymbol{\alpha}_{t-1}^{(j)})}. \quad (13)$$

Our approach uses MitISEM at each time t to construct the importance density $g_t(\tilde{\boldsymbol{\alpha}}_t^{(j)}|\boldsymbol{\alpha}_{t-1}^{(j)})$ around the target density $p(\mathbf{y}_t|\tilde{\boldsymbol{\alpha}}_t^{(j)})p(\tilde{\boldsymbol{\alpha}}_t^{(j)}|\boldsymbol{\alpha}_{t-1}^{(j)})$. This has two advantages: first, as shown in Hoogerheide et al. (2012), the MitISEM proposal is constructed using mixture of Student t around the candidate. This allows to handle easily complex posterior densities, e.g. multimodal densities. Second, because the proposal is constructed at each time point t around the target $p(\mathbf{y}_t|\tilde{\boldsymbol{\alpha}}_t^{(j)})p(\tilde{\boldsymbol{\alpha}}_t^{(j)}|\boldsymbol{\alpha}_{t-1}^{(j)})$ that explicit consider the new observation y_t there is such phenomenon as particle depletion and the resampling step is not longer required.

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

- 1) **Initialization.** Draw $\tilde{\boldsymbol{\alpha}}_0^{(j)} \sim p(\boldsymbol{\alpha}_0)$ for $j = 1, \dots, M$.
- 2) **Recursion.** For $t = 1, \dots, T$ construct the candidate density $\tilde{g}_t(\tilde{\boldsymbol{\alpha}}_t^{(j)}|\boldsymbol{\alpha}_{t-1}^{(j)})$ using the MitISEM procedure that 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 degrees of freedom). Using the target density:
$$p(\mathbf{y}_t|\tilde{\boldsymbol{\alpha}}_t^{(j)})p(\tilde{\boldsymbol{\alpha}}_t^{(j)}|\boldsymbol{\alpha}_{t-1}^{(j)}),$$
update the mode and scale of the candidate distribution using the IS weighted EM algorithm.
 - b.) **Adaptation:** Improve the candidate density using the MitISEM procedure, see Appendix B.
- 3) **Draws.** Draws $\tilde{\boldsymbol{\alpha}}_t^{(j)}$ from the constructed density $\tilde{g}_t(\tilde{\boldsymbol{\alpha}}_t^{(j)}|\boldsymbol{\alpha}_{t-1}^{(j)})$ and approximate $E[h_t(\boldsymbol{\alpha}_t)|\mathbf{y}_{1:T}]$ by:

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

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

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

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

Here we present some 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, APF and M-Filter we use $M = 50,000$ particles. In the M-Filter the particles correspond to draws from the proposal density.

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

$$\begin{aligned} y_t &= \alpha_t + \varepsilon_t & \varepsilon_t &\sim \text{N}(0, \sigma_\varepsilon^2), \\ \alpha_t &= \alpha_{t-1} + \eta_t & \eta_t &\sim \text{N}(0, \sigma_\eta^2). \end{aligned} \tag{14}$$

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 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 local level model in (14) 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 1 reports the results for the model in equation (14). KF filter is the best filter, as expected, in terms of the minimum bias and the smallest computing time. The results of the non-linear filters, however, are in line with those of KF in terms of the bias measures. 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 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) given by:

$$\begin{aligned} y_t &= e^{(\alpha_t/2)} \varepsilon_t & \varepsilon_t &\sim \text{N}(0, \sigma_\varepsilon^2), \\ \alpha_t &= \mu + \phi \alpha_{t-1} + \eta_t & \eta_t &\sim \text{N}(0, \sigma_\eta^2), \end{aligned} \tag{15}$$

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 to 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

²We 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.

$h = \bar{\sigma}^2 \exp(\alpha_t)$ defined as:

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

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 closed to a constant.

Table 2 reports the results for the model in equation (15). In all 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 bias and estimation variability. In this model the computational speed is comparable between the three non-linear filters, namely PF, APF and M-Filter.

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

$$\begin{aligned} \mathbf{y}_t &= \Lambda \mathbf{f}_t + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t &\sim \text{N}(0, \Sigma), \\ \mathbf{f}_t &= \phi_1 \mathbf{f}_{t-1} + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim \text{N}(0, \mathbf{Q}). \end{aligned} \tag{16}$$

In this model the KF is the optimal filter as for the model in (14) and we use KF results as the benchmark case in order to compare the non-linear filters. Table 3 reports the results for the model in equation (16) for 100 Monte Carlo replication, $N = 20$ series and 2, 4, 6 and 10 factors. Due to the linear and Gaussian model structure in (16), KF leads to the best results in terms of the speed and accuracy, but the non-linear filter results are in line with those of KF. The M-Filter performs better than 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. In terms of the computing time, computing time in all filters increase with the number of factors. In all cases, however, M-Filter requires a shorter computing time compared to the other non-linear and non-Gaussian filters. Based on the set of simulation studies, we conclude that the proposed M-Filter has advantages in terms of reduced bias and computing time compared to the existing non-linear and non-Gaussian filters, namely PF and APF.

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, the Model-based Momentum (M.M.) and the Residual Momentum (R.M.) strategy, presented in Section 3.

Table 1: Monte Carlo results for $I=100$ replications of the linear and Gaussian model of equation (14) with $T=100$. Kalman Filter (KF), Bootstrap Particle Filter (PF), Auxiliary Particle Filter (APF) and MitISEM Filter (M-Filter) with 50,000 particles. The table reports Likelihood Bias (LB) with respect to KF. Absolute deviation defined as $Bias = 1/I \sum_{i=1}^I abs(\tilde{\alpha}_{t,i} - \alpha_{t,i})$ relative to the KF. The table also reports the variability defined as $Var = 1/I \sum_{i=1}^I (\tilde{\alpha}_{t,i} - \alpha_{t,i})^2$ 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	Bias	Var	LB	Bias	Var	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	Bias	Var	LB	Bias	Var	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

Table 2: Monte Carlo results for $I=100$ replications of the stochastic volatility model of equation (15) with $T=100$. Kalman Filter (KF), Bootstrap Particle Filter (PF), Auxiliary Particle Filter (APF) and MitISEM Filter (M-Filter) with 50,000 particles. The table the absolute deviation defined as $Bias = 1/I \sum_{i=1}^I abs(\tilde{\alpha}_{t,i} - \alpha_{t,i})$ a ratio to the KF. The table also report the variability defined as $Var = 1/I \sum_{i=1}^I (\tilde{\alpha}_{t,i} - \alpha_{t,i})^2$ as a ratio to the KF. The final column reports the computing time in seconds for the four filters.

CV	0.1		0.5		Time	
Model	Bias	Var	Bias	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	Bias	Var	Bias	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 3: Monte Carlo results for $I=100$ replications of the DFM with $T=100$ and $N = 20$. Kalman Filter (KF), Bootstrap Particle Filter (PF), Auxiliary Particle Filter (APF) and MitISEM Filter (M-Filter) with 50,000 particles. The table reports Likelihood Bias (LB) with respect to KF. Absolute deviation defined as $Bias = 1/I \sum_{i=1}^I abs(\tilde{\alpha}_{t,i} - \alpha_{t,i})$ relative to the KF. The table also reports the variability defined as $Var = 1/I \sum_{i=1}^I (\tilde{\alpha}_{t,i} - \alpha_{t,i})^2$ 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	Bias	Var	LB	Bias	Var	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	Bias	Var	LB	Bias	Var	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

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 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, Model and Residual Momentum, 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 models, Bayesian inference is performed with 5000 burn-in and 5000 posterior draws. 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). Thirdly, 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).

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 4. We report there means, volatilities, Sharpe Ratios and largest loss of each model and strategy.

We first focus on the mean returns in Table 4. The results differ substantially over alternative model and strategy combination. It is seen that the Model Momentum strategy gives poor results for simple VAR-N and DFM-N models. The complex model structures, DFM-SV2 and FAVAR-SV2 do not lead to better results or are worse than the results for the more basic models DFM-SV and FAVAR-SV for both momentum strategies. That is, the SV2 component leads mostly to over-fitting and 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 influences results strongly 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 Model Momentum strategy while using more simple model structures in combination with the Residual Momentum Strategy already leads to good results on mean returns. Apparently, using this latter strategy implies *a learning from past errors*.

We next compare the volatility of realized returns in Table 4. The differences be-

Table 4: 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. 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.

		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	

tween 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 residual momentum strategies. Given the same model class, model momentum 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 4, 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 4. 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 the a very low loss. For all models, except SV, the largest loss is substantially lower for the model momentum strategy compared to residual momentum. Clearly, choice of momentum strategy matters substantially for risk of returns.

So far we have focused on summarizing posterior measures. An important advantage of Bayesian inference is to obtain 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 residual momentum ($K = 3, L = 2$). 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 model momentum, has very high returns in some periods. Similarly, the best-performing strategies, DFM-SV, Residual Momentum, with $K = 3, L = 2$ and DFM-SV, Model Momentum, $K = 4, L = 1$ 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 Model Momentum 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. Residual Momentum leads to reasonable returns for the simple models, VAR-N, SV and DFM-N.

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

Table 5: 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.
<i>Mixture of basic models and two strategies</i>					
VAR-N & SV & DFM-N(4,2)	M.M. & R.M.	0.10 (0.01,0.18)	3.9 (3.6,4.2)	0.025 (0.002,0.047)	-23.0 (-28.8,-17.5)
<i>Mixture of strategies per model</i>					
VAR-N	M.M. & R.M.	0.09 (-0.03,0.20)	4.7 (4.0,4.5)	0.019 (-0.007,0.043)	-32.6 (-35.6,-20.9)
SV	M.M. & R.M.	0.13 (-0.02,0.28)	4.3 (3.9,4.6)	0.032 (-0.005,0.065)	-22.2 (-29.9,-16.1)
DFM-N(4,2)	M.M. & R.M.	0.03 (-0.12,0.17)	4.3 (4.0,4.7)	0.006 (-0.028,0.041)	-24.4 (-31.1,-16.8)

However, it also does not improve the results for the complex DFM-SV2 and FAVAR-SV2 models compared to their simpler cases. Using Model Momentum 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 Residual Momentum 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 behavior over time.

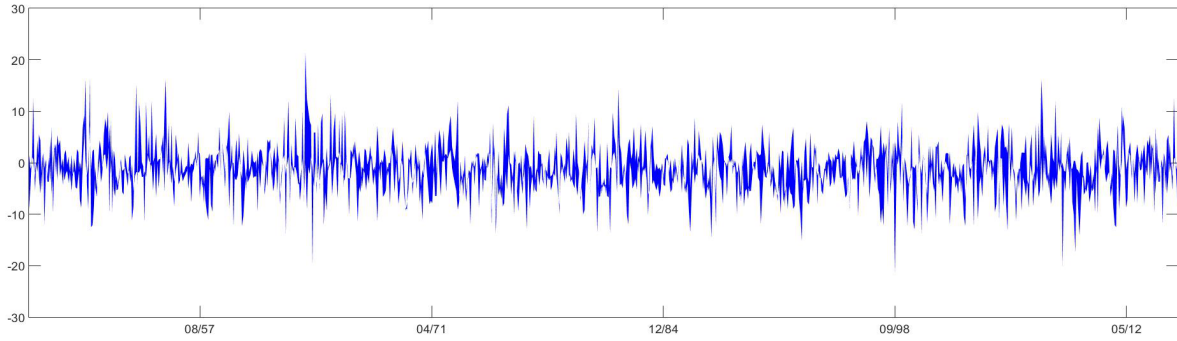
5.2 Combining possible mixtures of models and strategies

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.

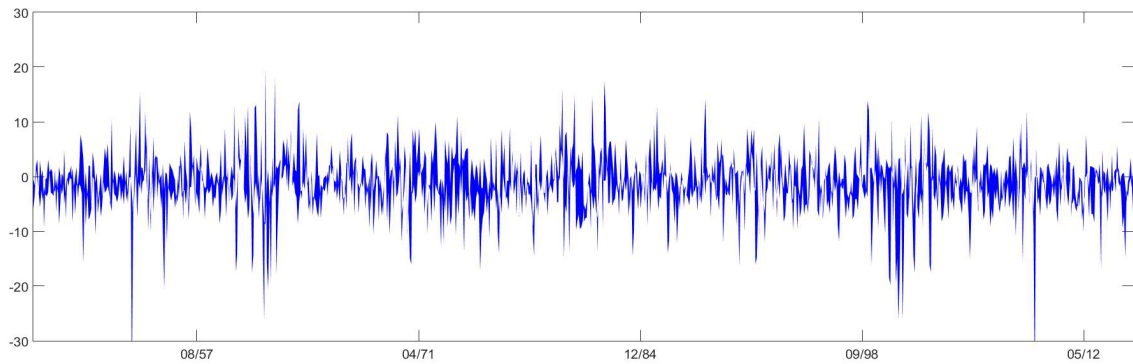
Combining basic model structures and investment strategies

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

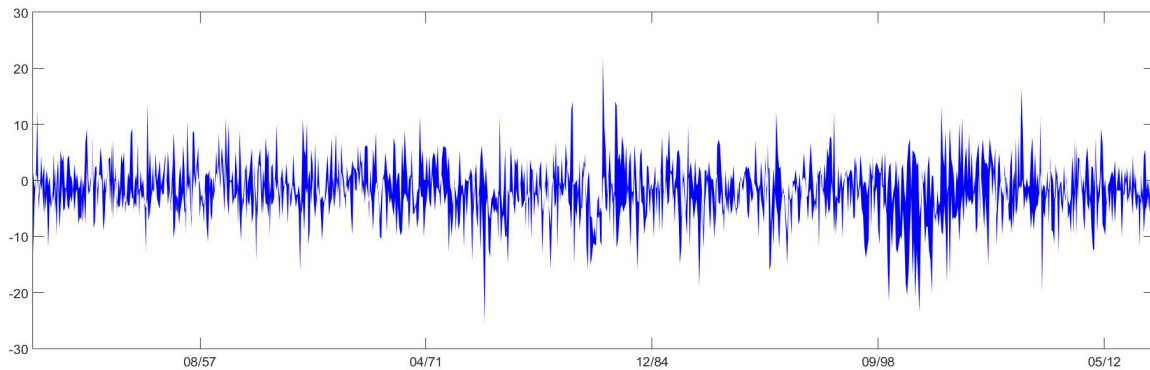
Figure 2: 99% realized return intervals for three selected models. DFM with $K = 1$ latent factor and $L = 1$ lags in Figure 2(a), DFM-SV with $K = 3$ latent factors and $L = 2$ lags in Figure 2(b) and DFM-SV with $K = 4$ latent factors and $L = 1$ lags in Figure 2(c).



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

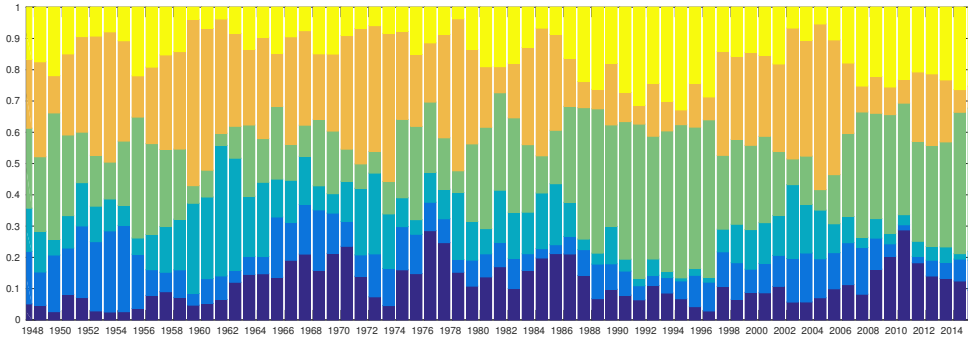


(b) DFM-SV, Residual Momentum, $K = 3$, $L = 2$

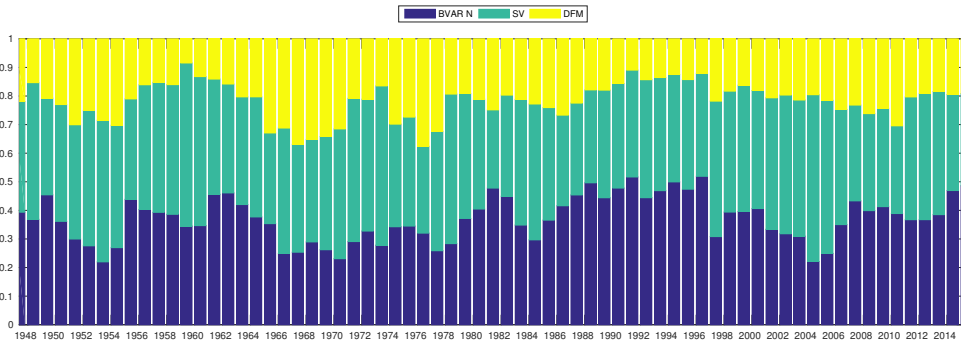


(c) DFM-SV, Model Momentum, $K = 4$, $L = 1$

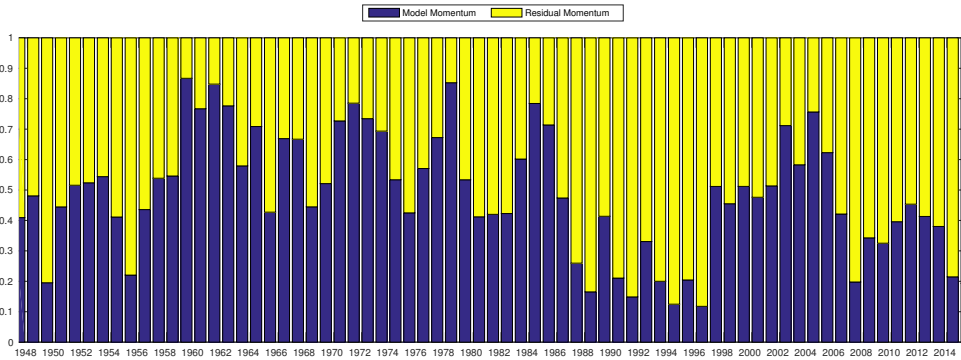
Figure 3: Posterior weights from the six-component mixture of three basic models (VAR-N, SV and DFM(4,2)) and two strategies (model momentum and residual momentum). The figure presents posterior means of all component weights in Figure 3(a), cumulative weights for each model in Figure 3(b) and cumulative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



(a) All Weights



(b) Cumulative Model Weights



(c) Cumulative Strategy Weights

pattern of the series into account, an 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, 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 (11):

$$f(r_t|I_{\mathbf{K}}) = \sum_{m=1}^3 \sum_{s=1}^2 w_{m,s,t} \int_{\mathbb{R}} f(r_t|\tilde{r}_{m,s,t}, I_k) f(\tilde{r}_{m,s,t}|I_k) d\tilde{r}_{m,s,t}, \quad (17)$$

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 model momentum and residual momentum.

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. Residual momentum 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 5. 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 5 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 5 leads to the following conclusions in terms of the *return and risk features*. The top panel of Table 5 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 than those of the individual model and strategy combinations in the top two panels of Table 4. 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 5 lead to better risk features compared to the individual components in Table 4, while model and strategy mixtures in the top panel of Table 5 lead in most cases to lower variances compared to the strategy mixtures in the bottom panel of Table 5.

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 4 or the mixture of strategies in the bottom panel of Table 5. 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 model momentum 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 the latter model we consider the case where one uses the density combination scheme to optimize over the number of factors from 1:4 and number of lags from 1:2. In Table 6 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 6, 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

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 $K = 1, \dots, 4$ 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.

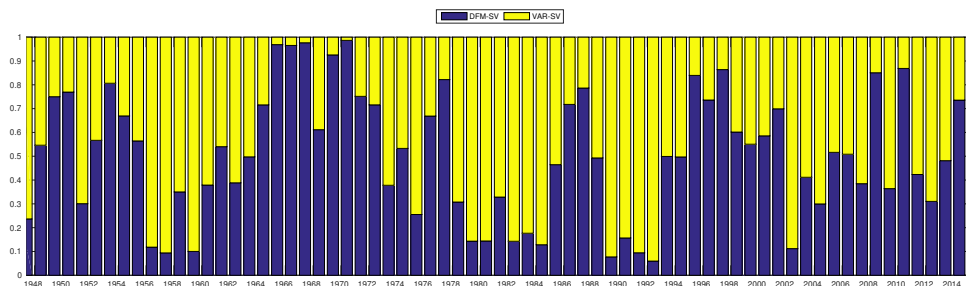


Table 6: Returns and risk from a mixture of two flexible models and two investment strategies. The top panel shows the return and risk from two model combinations (VAR-SV, DFM-SV(1:4, 1:2)) combined with two investment strategies (M.M. R.M.). The bottom panel reports these results for the mixture of 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.
<i>Mixture of two flexible models and strategies</i>					
VAR-SV & DFM-SV(1:4,1:2)	M.M. & R.M.	0.15 (0.08, 0.22)	3.7 (3.5, 3.9)	0.041 (0.021, 0.061)	-21.6 (-26.4, -16.4)
<i>Mixture of strategies per model</i>					
VAR-SV	M.M. & R.M.	0.23 (0.11, 0.35)	4.5 (4.2, 4.9)	0.051 (0.024, 0.080)	-37.2 (-37.3, -36.8)
DFM-SV(1:4,1:2)	M.M. & R.M.	0.06 (0.00, 0.12)	3.4 (3.2, 3.5)	0.018 (0.000, 0.036)	-14.4 (-20.1, -11.0)

Table 7: 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.

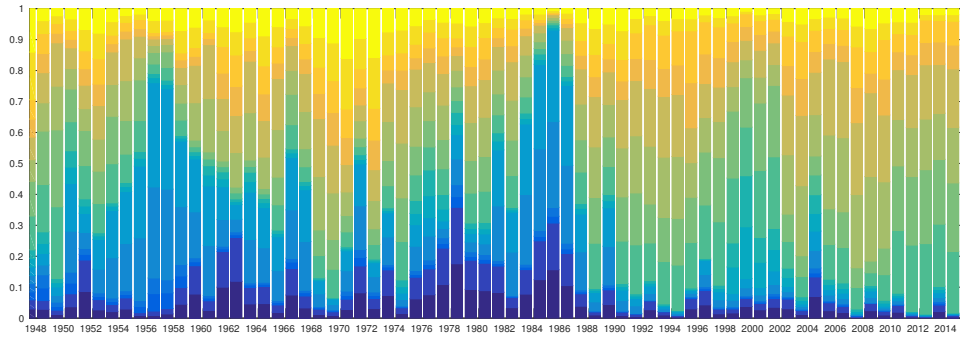
Model	Strategy	Mean	Vol.	S.R.	L.L.
<i>Mixture of basic models and two strategies</i>					
FAVAR-SV(1:4, 1:2)	M.M. & R.M.	0.18 (0.14, 0.22)	4.5 (4.5, 4.6)	0.039 (0.031, 0.048)	-34.8 (-35.0, -34.6)
<i>Mixture of strategies per model</i>					
FAVAR-SV(1, 1)	M.M. & R.M.	0.11 (0.02, 0.19)	4.5 (4.4, 4.6)	0.024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

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.

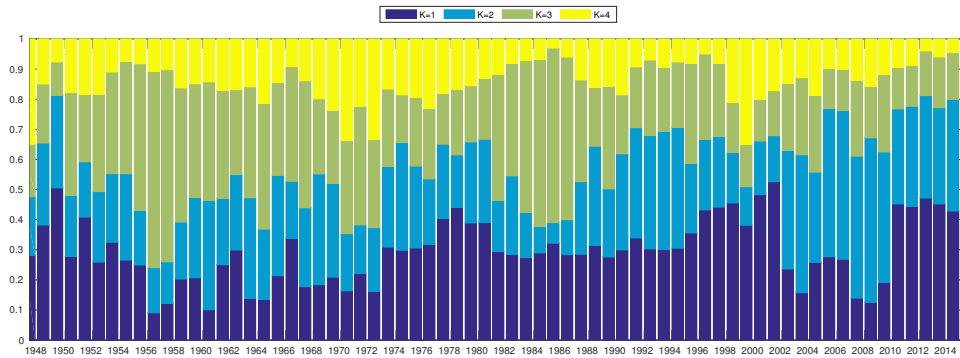
Combining a very flexible parametric model and a mixture of investment strategies

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

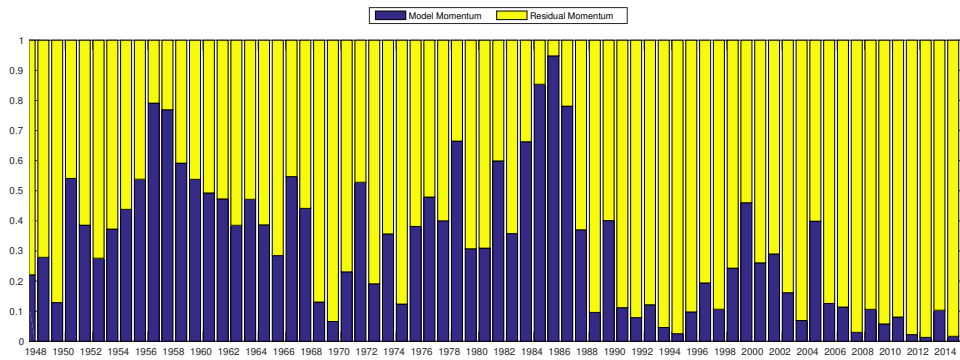
Figure 5: Posterior weights from DECO combination of 16 FAVAR-SV model and strategy combinations: all model weights, cumulative weights of each strategy and cumulative weights of number of factors in the model. The figure presents posterior means of all component weights in Figure 5(a), cumulative weights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

scheme in this case, in relation to the general equation (11), is as follows:

$$f(r_{kt}|I_{\mathbf{K}}) = \sum_{m=1}^4 \sum_{l=1}^2 \sum_{s=1}^2 w_{m,l,s,t} \int_{\mathbb{R}} f(r_t|\tilde{r}_{m,l,s,t}, I_k) f(\tilde{r}_{m,l,s,t}|I_k) d\tilde{r}_{m,l,s,t}, \quad (18)$$

where all combined models are member of that FAVAR-SV class and $m = 1, \dots, 4$ 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 model momentum and residual momentum 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. A single one of the considered model specifications or investment strategies do not seem 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 recent period 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, 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. Model based momentum appears to be important until the 1990s while residual momentum is important 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 7. 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 4. By comparing the results from Table 4 with the bottom panel of Table 7, 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 presented a dynamic asset-allocation approach which combines alternative models and momentum strategies that updates the portfolios as well as portfolio strategies at every decision period. A set of flexible dynamic models with different number of latent factors, a vector autoregressive and a stochastic volatility component is specified to capture short and long-run dynamics. Next, this set of models is combined with a set of portfolio strategies. The density combination scheme in [Casarin et al. \(2016\)](#) is extended to obtain mixtures of model structures and momentum strategies at each time period. In addition, a new and efficient non-linear and non-Gaussian filter is introduced, based on the MitISEM approach, see [Hoogerheide et al. \(2012\)](#), and it is shown that the proposed filter leads to substantial accuracy and speed gains. Using US industry portfolios for the period 1926-2015 as data, our results indicate that time-varying combinations of flexible components of the class of FAVAR-SV models and two strategies lead to better return and risk features than very simple and very complex models. More specifically, flexible 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 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. 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 the different model structures used in Section 2 in relation to the general model formulation in equation (1).

Linear and Gaussian Dynamic Factor Model (DFM)

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

$$\begin{aligned} \mathbf{y}_t &= \Lambda \mathbf{f}_t + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t &\sim \text{N}(0, \Sigma), \\ \mathbf{f}_t &= \phi_1 \mathbf{f}_{t-1} + \dots + \phi_L \mathbf{f}_{t-L} + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim \text{N}(0, \mathbf{Q}), \end{aligned} \quad (19)$$

that is a linear and Gaussian DFM. To estimate this model we assume the following priors:

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

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

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

- 2) The loading parameters has normal priors, $\underline{\Lambda} \sim \text{N}(\underline{\mu}, \underline{\mathbf{C}})$ where $\underline{\mu} = 0$ and $\underline{\mathbf{C}} = \mathbf{I}$;
- 3) The prior for the autoregressive parameters $\Phi = [\phi_1, \dots, \phi_L]$ and latent errors variance \mathbf{Q} are diffuse conjugate Normal-Wishart prior:

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

where $\underline{\Phi} = \text{vec}(\Phi)$ is the Φ stacked in a column vector of length $L \times K^2$, where L are the number of lags in the latent factor. 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 \mathbf{Q}_0 are set to the residual variances of the corresponding univariate autoregressions, $\hat{\sigma}_{\eta, kk}^2$ for $k = 1, \dots, K$. The Ω_0 diagonal elements are set on k lagged j 'th variable in i 'th equation equals $\sigma_i^2 / k \sigma_j^2$.

Defining $\Lambda_i = (\lambda_{i,1}, \dots, \lambda_{i,k})$ for $i = 1, \dots, N$ 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:

$$\bar{\sigma}_{\varepsilon, ii}^2 \sim \text{IG}\left(\frac{v + T}{2}, \frac{vs^2 + d_i}{2}\right)$$

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

- 2) The draws of the loading parameters taking into consideration the restrictions are given by:

- a) For $i = 1, \dots, K$, draw $\bar{\Lambda}_i \sim N(\bar{m}_i, \bar{C}_i)\mathbf{I}(\lambda_{ii} > 0)$ where $\bar{m}_i = \bar{C}_i(\underline{C}_i^{-1}\underline{\mu}_i + \sigma_{ii}^{-2}f_i'y_i)$ and $\bar{C}_i^{-1} = \underline{C}_i^{-1} + \sigma_{ii}^{-2}f_i'f_i$
 - b) For $i = K+1, \dots, N$ draw $\bar{\Lambda}_i \sim N(\bar{m}_i, \bar{C}_i)$ where $\bar{m}_i = \bar{C}_i(\underline{C}_i^{-1}\underline{\mu}_i + \sigma_{ii}^{-2}f_i'y_i)$ and $\bar{C}_i^{-1} = \underline{C}_i^{-1} + \sigma_{ii}^{-2}f_i'f_i$
- 3) Regarding the $\bar{\Phi}$ and \mathbf{Q} we have a standard VAR form that can be estimated equation by equation, to obtain the following posterior:
- a) Draws the $\bar{\mathbf{Q}}$ from the iW($\hat{\mathbf{Q}}, T + K + N + 2$) where $\hat{\mathbf{Q}} = \underline{\mathbf{Q}} + \hat{\Gamma}'\hat{\Gamma} + \hat{\Phi}'[\Omega_0 + (\hat{\Gamma}_t'\hat{\Gamma}_t)^{-1}]^{-1}\hat{\Phi}$ and $\hat{\Gamma}$ is the matrix of OLS residuals.
 - b) Draw the $\bar{\Phi}$ from the conditional normal according to:
- $$\bar{\Phi} \sim N(\text{vec}(\tilde{\Phi}), \mathbf{Q} \otimes \tilde{\Omega}) \quad (20)$$
- 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 dynamic factor model with stochastic volatility in the idiosyncratic errors by setting $\beta = 0$ in (1) and specifying the error covariances as follows:

$$\begin{aligned} \mathbf{y}_t &= \Lambda \mathbf{f}_t + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t &\sim N(0, \Sigma_t), \\ \mathbf{f}_t &= \phi_1 \mathbf{f}_{t-1} + \dots + \phi_L \mathbf{f}_{t-L} + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim N(0, \mathbf{Q}), \end{aligned} \quad (21)$$

where we specify the following time varying variance matrix:

$$\Sigma_t = \begin{pmatrix} \sigma_{11,t}^2 & 0 & \dots & 0 \\ 0 & \sigma_{22,t}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{ii,t}^2 \end{pmatrix}, i = 1, \dots, N, \quad (22)$$

then we have the DFM-SV model. We assume that the log volatilities $h_{it} = \log(\sigma_{ii,t}^2)$ follows a stationary and mean reverting process:

$$\mathbf{h}_t = \boldsymbol{\mu} + \boldsymbol{\psi} \mathbf{h}_{t-1} + \boldsymbol{\zeta}_t, \quad \boldsymbol{\zeta}_t \sim N(0, \Gamma), \quad \psi_i \in (-1, 1).$$

Starting from equation (21) and taking the difference $\boldsymbol{\varepsilon}_t = \mathbf{y}_t - \Lambda \mathbf{f}_t = \mathbf{y}_t^*$, and then the squares plus an offset constant we get:

$$\begin{aligned} \mathbf{y}_t^{**} &= \log[(\mathbf{y}_t^*)^2 + \bar{c}], \\ \mathbf{y}_t^{**} &= 2\mathbf{h}_t + \mathbf{e}_t, \\ \mathbf{h}_t &= \mu + \psi \mathbf{h}_{t-1} + \xi_t, \quad \xi_t \sim \text{N}(0, \Delta), \end{aligned} \tag{23}$$

where $\mathbf{e}_t = \log(\boldsymbol{\varepsilon}_t)$ is distributed as a $\chi^2(1)$. This prevent the use of the standard Kalman filter recursions and simulations algorithms as in [Carter and Kohn \(1994\)](#). To solve this problem [Kim et al. \(1998\)](#), employ a data augmentation approach, introduce a new state variable $\mathbf{s}_T = \{\mathbf{s}_1, \dots, \mathbf{s}_t\}$ that allows to rewrite the linear non-Gaussian state space model in equation (23) as a conditional linear Gaussian state space form. This is due to the following approximation:

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

where τ_j , ν_j^2 and q_j for $j = 1, \dots, 7$ are constant specified in [Kim et al. \(1998\)](#). The draw of the error \mathbf{e}_t implies:

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

and draws from the sequence of states \mathbf{s}_T can be obtained using:

$$\Pr(\mathbf{s}_t = j | \mathbf{y}_t^{**}, \mathbf{h}_t) \propto q_j f_{\text{N}}(\mathbf{y}_t^{**} | 2\mathbf{h}_t + \tau_j - 1.2704, \nu_j^2), \tag{24}$$

where $f_{\text{N}}(\cdot)$ denotes the kernel of a normal density and $j = 1, \dots, 7$, $t = 1, \dots, T$. Conditional on \mathbf{s}_T the model is linear and Gaussian and the algorithm of [Carter and Kohn \(1994\)](#) can be used. The priors remains as the one described before with the difference in the SV parameters: μ , ψ and variance of the errors Γ .

We choose priors for μ_i , ψ_i , the initial log volatility h_{i1} and γ_{ii}^{-2} from the normal-gamma family:

$$\begin{aligned} \left[\begin{pmatrix} \mu_i \\ \psi_i \end{pmatrix} \right] &\sim \text{N} \left[\begin{pmatrix} \underline{\mathbf{m}}_{\mu_i} \\ \underline{\mathbf{m}}_{\psi_i} \end{pmatrix}, \begin{pmatrix} \underline{\mathbf{V}}_{\mu_i} & 0 \\ 0 & \underline{\mathbf{V}}_{\psi_i} \end{pmatrix} \right], \quad \psi_i \in (-1, 1), \\ \gamma_{ii}^{-2} &\sim \text{G}(1/\underline{\mathbf{k}}_{\gamma}, 1). \end{aligned}$$

Following [Pettenuzzo and Ravazzolo \(2016\)](#), we set $\underline{\mathbf{k}}_{\gamma} = 0.01$, $\underline{\mathbf{m}}_{\mu_i} = 0$, $\underline{\mathbf{m}}_{\psi_i} = 0.95$, $\underline{\mathbf{V}}_{\mu_i} = 10$ and $\underline{\mathbf{V}}_{\psi_i} = 1.0e^{-06}$. These prior values imply a strong autocorrelation structure for the h_{it} . We keep the priors on the remaining parameters as described in the earlier section.

For this model, the Gibbs sampling scheme is as follows:

- 1) Initialize $\mathbf{f}_t^{(0)}$, $\mathbf{h}_t^{(0)}$, $\Lambda^{(0)}$, $\mathbf{Q}^{(0)}$.

- 2) Draws the latent states $p(\mathbf{f}_t | \Lambda, \mathbf{Q}, \Sigma_t, \mathbf{h}_t)$ using the FF-BS algorithm described in [Carter and Kohn \(1994\)](#).
- 3) Conditional on \mathbf{h}_t and Λ draw the indicator variable \mathbf{s}_T for the mixture accordingly to equation (24).
- 4) Draws the sequence of stochastic volatilities $p(\mathbf{h}_t | \Lambda, \mathbf{f}_t, \mathbf{s}_t, \mu, \psi)$ from the conditional linear and Gaussian system using [Carter and Kohn \(1994\)](#).
- 5) Draws the stochastic volatility variances $p(\gamma_{ii}^2 | \mathbf{h}_t, \mu, \psi)$ from the following posterior:

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

- 6) Draw the SV parameters jointly:

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

where

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

and

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

- 7) Go to step 2.

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

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

$$\begin{aligned} \mathbf{y}_t &= \Lambda \mathbf{f}_t + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t &\sim \text{N}(0, \Sigma_t), \\ \mathbf{f}_t &= \phi_1 \mathbf{f}_{t-1} + \dots + \phi_L \mathbf{f}_{t-L} + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim \text{N}(0, \mathbf{Q}_t), \end{aligned} \quad (25)$$

with the idiosyncratic errors defined as in equation (22) and latent error variances given by:

$$\mathbf{Q}_t = \begin{pmatrix} \eta_{11,t}^2 & 0 & \dots & 0 \\ 0 & \eta_{22,t}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \eta_{KK,t}^2 \end{pmatrix}, i = 1, \dots, K, \quad (26)$$

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

$$\mathbf{k}_t = \omega + \beta \mathbf{k}_{t-1} + \xi_t, \quad \xi_t \sim N(0, \Xi).$$

The estimation of this model proceeds as before with an added step in the Gibbs Sampler. This step extract the later variance processes extending the procedure described before.

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

Assuming in equation (1) $\beta \neq 0$ and a time varying variance-covariance matrix for the idiosyncratic and latent errors we obtain the following model:

$$\begin{aligned} \mathbf{y}_t &= \beta \mathbf{x}_t + \Lambda \mathbf{f}_t + \boldsymbol{\varepsilon}_t & \boldsymbol{\varepsilon}_t &\sim N(0, \Sigma_t), \\ \mathbf{f}_t &= \phi_1 \mathbf{f}_{t-1} + \dots + \phi_L \mathbf{f}_{t-L} + \boldsymbol{\eta}_t & \boldsymbol{\eta}_t &\sim N(0, Q_t). \end{aligned} \quad (27)$$

The FAVAR model extends the state equation by defining \mathbf{x}_t as the vector of lagged dependent variables. This leads to a VAR form in particular in the state equation of (27):

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

see also [Stock and Watson \(2005\)](#). Conditional on the latent states, the estimation of the VAR parameters β are similar to those of univariate linear regression models, hence Bayesian inference is standard. The two proposed FAVAR models are defined by a stochastic volatility component in the idiosyncratic disturbances (FAVAR-SV) and stochastic volatility components in the idiosyncratic and latent disturbances (FAVAR-SV). We refer to the earlier sections of this appendix for the inference of the SV components given the rest of the parameters.

Appendix B

In this section, we present the details of the proposed M-Filter in Section 4. For a detailed discussion of the general MitISEM procedure, we refer to [Hoogerheide et al. \(2012\)](#). In the derivations we assume that the sets of estimated parameters $\hat{\boldsymbol{\theta}}$ are known and we omit them for the sake of notation. The M-Filter follows those recursions:

- 1) **Initialization.** Draw $\tilde{\boldsymbol{\alpha}}_0^{(j)} \sim p(\boldsymbol{\alpha}_0)$ for $j = 1, \dots, M$.
- 2) **Recursion.** For $t = 1, \dots, T$ construct the candidate density $\tilde{g}_t(\tilde{\boldsymbol{\alpha}}_t^{(j)} | \boldsymbol{\alpha}_{t-1}^{(j)})$ using the MitISEM procedure that can be summarize as follows:

- a.) **Initialization:** Simulate draws $\tilde{\alpha}_t^{(j)}$ from a ‘naive’ candidate distribution with density $g_n(\cdot)$. Using the target density:

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

update the mode and scale of the candidate distribution using the IS weighted EM algorithm.

- b.) **Adaptation:** Estimate the target distribution’s mean and covariance matrix using IS with the draws $\tilde{\alpha}_t^{(j)}$ from a naive distribution $g_n(\cdot)$ (e.g. a Student- t with v degrees of freedom). Use these estimates as the mode and scale matrix of Student- t adaptive density $\tilde{g}_a(\cdot)$. Draw a sample $\tilde{\alpha}_t^{(j)}$ from this adaptive Student- t distribution and compute the IS weights for this sample.
- c.) Apply the **IS-weighted EM algorithm** given the latest IS weights and the sample of step b). The output consists of the new candidate density with $H = 1$ components $\tilde{g}_t^H(\cdot)$ with optimized parameters. Draw a new sample $\tilde{\alpha}_t^{(j)}$ from this candidate and compute corresponding IS weights.
- d.) **Iterate on the number of mixture components.** Given a draw $\tilde{\alpha}_t^{(j)}$ and the corresponding IS weights:

$$\tilde{w}_t^{(j)} = \frac{p(\mathbf{y}_t | \tilde{\alpha}_t^{(j)})p(\tilde{\alpha}_t^{(j)} | \alpha_{t-1}^{(j)})}{\tilde{g}_t^H(\tilde{\alpha}_t^{(j)} | \alpha_{t-1}^{(j)})}$$

from the current mixture of H components construct the additional component in the mixture candidate density. This choice ensures that the new component covers a region of the target density in which the previous candidate mixture had relatively too little probability mass. Given the latest IS weights and the drawn sample from the current mixture of H components ($\tilde{g}_t^H(\cdot)$), apply the IS-weighted EM algorithm to optimize *each* mixture component parameters. Draw a new sample $\tilde{\alpha}_t^{(j)}$ from the mixture of $H + 1$ components ($\tilde{g}_t^{H+1}(\cdot)$) and compute corresponding IS weights.

- e.) **Assess convergence of the candidate density’s quality by inspecting the IS weights** and return to step d) unless the algorithm has converged.
- 3) **Draws.** Draws $\tilde{\alpha}_t^{(j)}$ from the constructed density $\tilde{g}_t^H(\tilde{\alpha}_t^{(j)} | \alpha_{t-1}^{(j)})$ and approximate $E[h_t(\alpha_t) | \mathbf{y}_{1:T}]$ by:

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

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

tion 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 in case of a complex model structure and a high 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.

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, Elias 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. ISSN 1548-7660.
URL <https://www.jstatsoft.org/index.php/jss/article/view/v068i03>
- 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.
- 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.
- 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.