Forecast Density Combinations of Dynamic Models and Data Driven Portfolio Strategies*

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

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

September 17, 2018

Abstract

Typical stylized facts about the time series pattern of monthly returns of ten US industries are analyzed over the period 1926M7–2015M6. This leads to the specification of a set of dynamic models with forecasts that are directly connected - without the specification of a utility or other scoring function - with a set of data driven portfolio strategies. The latter ones are based on the practice in financial investment to take advantage of a positive or negative momentum in industry returns. The resulting dynamic asset-allocation model is specified in probabilistic terms as a combination of return distributions resulting from multiple pairs of models and strategies and is represented as a nonlinear state space model. The combination weights are allowed to be cross-correlated and correlated over time, defined through the use of feedback mechanisms that enable learning. Diagnostic analysis of posterior residuals gives insight into model and strategy incompleteness or misspecification. Simulation-based Bayesian inference is performed in an efficient and robust manner by the introduction of a non-linear and non-Gaussian filter, labeled the M-Filter. In our empirical application a combination of a smaller set of flexible models performs better in terms of improved features of expected return and risk than a larger combination of basic model structures. Dynamic patterns in combination weights and diagnostic learning provide useful signals for improved modeling and policy, in particular, from a risk-management perspective.

*We thank two anonymous referees and the guest-editors, Sylvia Kaufman and Sylvia Frühwirth-Schnatter, for their valuable comments on earlier versions of this paper which led us to prepare an extended and substantially revised version of Baştürk, Grassi, Hoogerheide and Van Dijk (2016).

1 Introduction

Given several stylized facts about the time series pattern of monthly returns of ten US industries over an extensive period from 1926M7 - 2015M6, a set of dynamic models is specified based on these facts[1]. Forecasts from this model set are directly connected - without the specification of a utility or other scoring function - with a set of data driven portfolio strategies. These strategies refer to the basic practice in financial investment that one invests in the 'winner' industry and goes short in the 'loser industry', corresponding to the industries with the highest and lowest cumulative returns in past periods. That is, one aims to take advantage of a positive or a negative 'momentum' in returns of particular industries.

We show that this dynamic asset-allocation model can be represented in probabilistic terms as a density combination of model forecasts as well as strategy returns. The combination weights are allowed to be cross-correlated and correlated over time where the latter are defined through the use of feedback mechanisms that enable learning. Our approach extends the mixture of experts analysis in Jacobs et al. (1991); Jordan and Jacobs (1994); Jordan and Xu (1995); Peng et al. (1996). Further, we allow for model and strategy incompleteness. This enables us to study misspecification effects through diagnostic analysis of economic results and posterior residuals. This, to the best of our knowledge, novel methodology provides dynamic asset-allocations using a learning period for optimal weights at every decision period.

To achieve this, we present an extension of the Forecast Density Combination (FDC) scheme from Billio et al. (2013) to include sets of model forecasts and strategy returns. Using this scheme in a fully Bayesian setting, another novel contribution of our approach is that the policy recommendation to an investor about different portfolio scenarios includes the uncertainty in the returns. This is important from a risk management perspective. With merely a standard point forecast, an investor has no information on e.g. the Value-at-Risk of his/her portfolio.

For the numerical evaluation of the densities involved we make use of the result that this FDC can be represented as a non-linear state space model. Inference on density features brings a challenge in terms of estimation efficiency and robustness and amount of computing time, particularly in case of a large number of models and strategies. In order to tackle this, we introduce a novel non-linear and non-Gaussian filter, labeled the M-Filter, which is embedded in the density combination procedure. This filter is based on the MitISEM procedure recently proposed by Hoogerheide et al. (2012) and further developed in Baştürk et al. (2016) and Baştürk et al. (2017).

[1]These industry returns are constructed by equally weighting all stock returns in the specific industry, which is similar to Moskowitz and Grinblatt (1999). The data are retrieved from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french on 24/10/2015.

We investigate the performance of the FDC method using data on ten US industry returns over the period 1926M7 - 2015M6. The results of our empirical analysis contain valuable information for further research as well as informative signals about the scenarios of alternative portfolio policies. This may be useful information for a large financial investment firm, like a pension fund. The results refer to three central issues. First, we obtain evidence that averaging over density combinations of sets of model forecasts and strategy returns pays off in terms of expected return and risk features. The forecasts of the model sets help to improve expected return while the strategy sets help to reduce risk features. Basic model structures and strategies with fixed weights perform worse in terms of expected return and Sharpe ratio. Second, we obtain evidence that the dynamic patterns of the weights in these combinations differ in quiet and more volatile periods. Basic learning mechanisms for the weights are useful instruments in this respect. Third, there exist adverse effects of misspecification of the model and strategy set on the results. Diagnostic learning about economic information and about posterior residual patterns is helpful for improved modeling and policy. We emphasize that our empirical results are conditional upon an information set which consists of our data set, US industrial portfolios over the period between 1926M7 and 2015M6, and our specified model and strategy set.

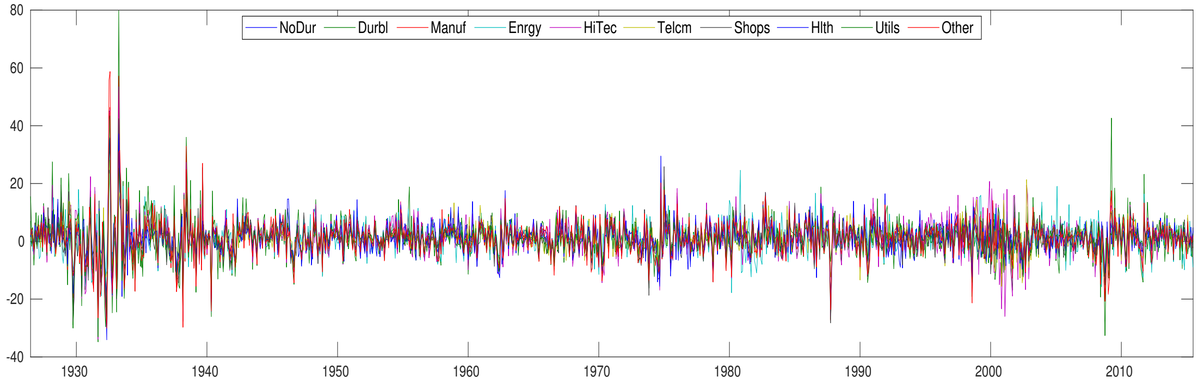
The contents of this paper are structured as follows: Section 2 introduces the dynamic models used for US industry returns. Section 3 describes the direct connection between forecasts of sets of dynamic models and returns of portfolio strategies. Section 4 covers the extended FDC scheme and introduces the M-Filter. Section 5 contains the empirical application using returns from ten US industries. Section 6 concludes. An on-line Appendix contains additional results.

2 Stylized facts about ten US industry returns leading to dynamic models

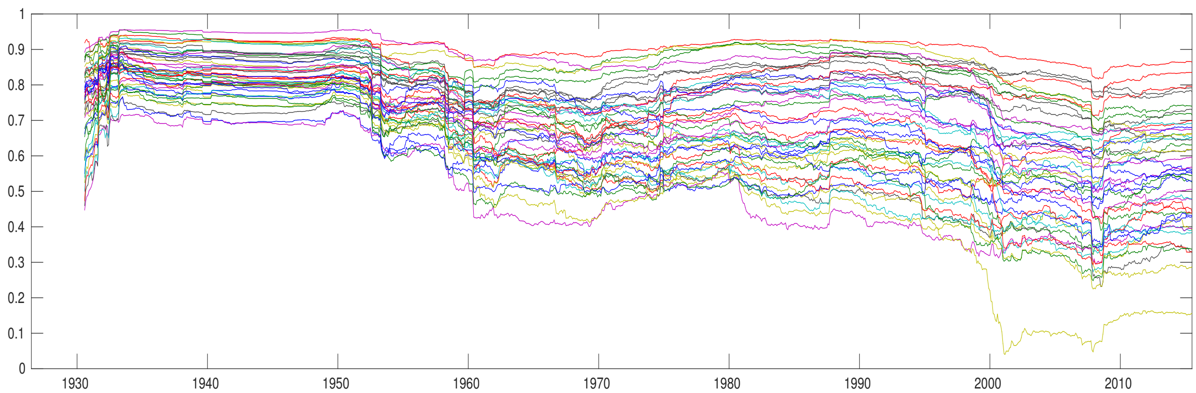
Traditional factor models rely on macro or firm specific factors to explain expected payoffs of financial assets, see Fama and French (1992, 1993, 2015). In the literature, several dynamic factor models, with different long and short-run dynamics for returns, are shown to be useful in capturing such data properties, see Ng et al. (1992), Quintana et al. (1995), Aguilar and West (2000) and Han (2006) among several others. These models are components of the Factor-Augmented Vector AutoRegressive model (FAVAR), see Bernanke et al. (2005) and Stock and Watson (2005). Members of this class of models are applied for portfolio construction in Aguilar and West (2000), Talih and Hengartner (2005), Engle and Colacito (2006), Carvalho et al. (2011) and Zhou et al. (2014), among many others.

In this section we summarize several stylized facts about the time series pattern of monthly returns of ten US industries over the period, 1926M7 - 2015M6. This leads in a relatively

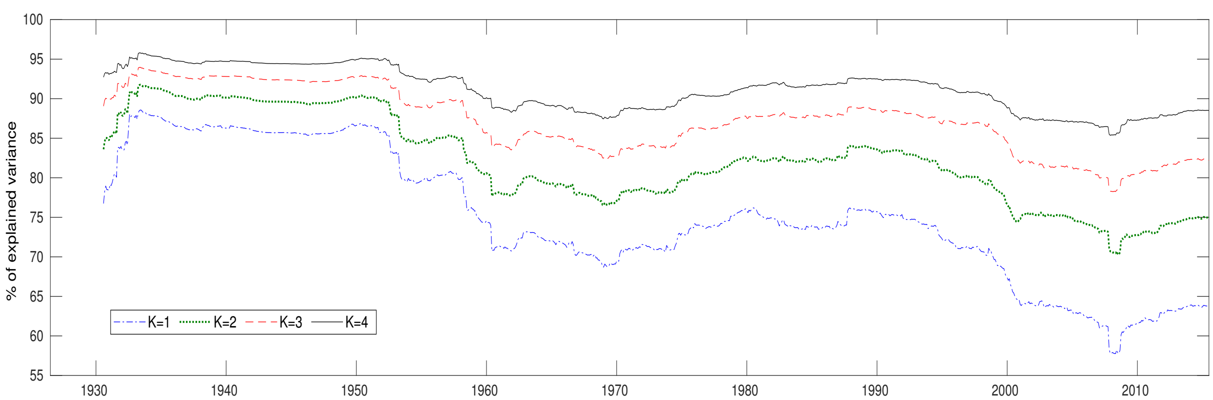
Figure 1: *Monthly percentage returns (a), explained variation of principle components (b), and canonical correlations (c), across 10 US industry portfolios 1926M7-2015M6. Industry portfolios are abbreviated as follows: 'NoDur for non-durables', Durbl for 'durables', Manuf for 'manufacturing',Enrgy for 'energy', HiTec for 'hi-tech', Telcm for 'telecom', 'Shops', Hlth for 'health', Utils for 'utilities' and the final category 'Others'.*



(a) Monthly percentage returns



(b) Canonical correlations



(c) Percentage of explained variation by PCA natural way to the specification of our set of dynamic models.

Figure 1(a) presents monthly returns of ten industries where the industries are abbreviated as follows: 'NoDur' for consumer non-durables (food, tobacco, textiles, apparel, leather, toys), 'Durbl' for consumer durables (cars, TV's, furniture, household appliances), 'Manuf' for manufacturing (machinery, trucks, planes, chemicals, office furniture, paper, commercial printing), 'Enrgy' for oil, gas, and coal extraction and products, 'HiTec' for business equipment (computers, software, and electronic equipment), 'Telcm' for telephone and television transmission, 'Shops' for wholesale, retail, and some services (laundries, repair shops), 'Hlth' for health care, medical equipment, and drugs, 'Utils' for utilities, 'Other' for other industries. Next, 45 pairwise correlations of the 10 industry returns in Figure 1(b) and 4 principal components in Figure 1(c) are based on moving windows with 240 monthly observations. We use the first 50 observations as the initial sample and expand calculation windows until observation 240.

One may observe, at least, four stylized facts from Figures 1(a)–1(c): In the top figure, a stationary autoregressive time series pattern is seen for all return series with, in addition, clear volatility clustering also common to all series. Strong cross-section correlations between returns with a time-varying pattern are shown in the middle figure, and fourthly, the bottom figure indicates that the total variation in the series is well captured with one to four principal components. We emphasize that the explained variation of these components is time-varying.

Given these typical data features, we consider several dynamic models with clearly distinct short and long-run dynamics and different features of the disturbance distributions. All models considered are members or combinations of members of the class of Factor Augmented Vector AutoRegressive models extended to include Stochastic Volatility of the idiosyncratic disturbances (FAVAR-SV) :

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

where the dependent variable $\mathbf{y}_t = (y_{1,t}, \ldots, y_{N,t})'$ is the $N \times 1$ vector of industrial portfolio returns, where $y_{i,t}$ denotes the return from industry i at time t and the time series runs from $t = 1, \ldots, T$. The $C \times 1$ vector of predetermined variables $\mathbf{x}_t$ may contain explanatory variables as well as lagged dependent variables. The $K \times 1$ vector $\mathbf{f}_t$ contains unobservable factors, where ϕ_j for $j = 1, \ldots L$ is a $K \times K$ matrix of autoregressive coefficients at lag j. Λ is an $N \times K$ matrix of factor loadings. In addition we define a time-varying variance-covariance matrix for the idiosyncratic disturbances, Σ_t, and a fixed covariance matrix for the factor disturbances, Q. In all specifications Σ_t is a diagonal matrix.[2]

[2]We have also estimated models with a $t-$distribution and/or a time varying covariance matrix, Q_t. Both extensions led to overfitting and poor empirical and forecasting results. We have therefore deleted these models from our analysis. Particularly for the latter case, we acknowledge that the MCMC sampler

Different short and long-run dynamic behavior of member models of the FAVAR-SV class is obtained by specifying different assumptions regarding the predetermined variables $\mathbf{x}_t$, the factor structure $\mathbf{f}_t$, the idiosyncratic and factor disturbances. The basic dynamic factor model, denoted by DFM, assumes $\boldsymbol{\beta} = 0_{(N \times C)}$, a normal distribution for the idiosyncratic and factor disturbances with time-invariant variance-covariance matrices. Another basic model is the vector autoregressive model, denoted by VAR, and it is obtained by letting $\Lambda = 0_{(N \times K)}$, defining $\mathbf{x}_t$ as the lagged dependent variable and a time-invariant variance-covariance matrix of the disturbances. A third basic model, denoted by SV, has a stochastic volatility component in the idiosyncratic disturbances and $\boldsymbol{\beta} = 0$ and $\Lambda = 0$. We provide more details on the specification of the models in the on-line Appendix A together with their prior specification and Bayesian estimation procedures.
In our empirical analysis, reported in Section 5, we compare the performance of alternative combinations of models for forecasting and portfolio analysis. We start with exploring the contribution of each of the three basic models, VAR, SV and DFM, separately and as a model combination. As a next step we investigate combinations of more flexible models like VAR-SV and DFM-SV and finally, the general class of FAVAR-SV is investigated.

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

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

for any $K \times K$ invertible matrix R, which has K^2 free parameters. Hence at least K^2 restrictions are needed for the model to be identified, see Geweke and Zhou (1996), Lopes and West (2004), Bai and Peng (2015) and Frühwirth-Schnatter and Lopes (2018). In all models, we follow the identification scheme in Lopes and West (2004) and assume diagonal covariance matrices. See Chan et al. (2018) and Kaufmann and Schumacher (2017) for more recent specifications of identification in this class of models.

3 Connecting dynamic model forecasts directly with data driven portfolio strategies

Standard portfolio analysis compares realized returns from different portfolio strategies and selects the best performing one, see e.g. Aguilar and West (2000). But return forecasts using dynamic models do not lead directly to a practical policy tool for investors, that is, to a decision which portfolio strategy to follow. 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).

can be improved, see Kastner et al. (2017).

A novel contribution of this paper is to connect forecasts from the set of dynamic models of Section 2 directly with a set of data driven portfolio strategies without the need to specify a separate scoring function like a utility or loss function. Such portfolio strategies have also been proposed by Garlappi et al. (2006) and DeMiguel et al. (2007). Our approach differs from this literature since we consider sets of models as well as strategies and we make use of a Bayesian approach.

Standard Momentum (S.M.): As a benchmark data driven portfolio strategy, we consider so-called standard industry momentum. This does not involve a model structure but directly makes use of typical momentum patterns in the time series of monthly returns of our set of ten US industries. The practice is that one invests in the 'winner' industry and goes short in the 'loser industry', corresponding to the industries with the highest and lowest cumulative returns, say, in the past 12 periods. The selected momentum breakpoints correspond to, say, 90% and 10% quantiles for 10 industries, and these values can be adjusted for alternative momentum strategies. The economic intuition of this strategy is to capture market trends in industry returns.

Next, we list two portfolio strategies, based on the concept of momentum strategy, which are directly connected with in-sample forecasts from a model or a set of models. We note that our approach can be generalized to a wider selection of model-based portfolio strategies, such as those analyses in Gruber and West (2017).

Model based momentum (M.M.): To construct a portfolio based on this strategy, we use the fitted industry returns in the past period from one of the models or model sets of Section 2, go long in the industry with the highest fitted returns and go short in the industry with the lowest fitted returns. With ten industries, this corresponds to 90% and 10% quantiles of fitted returns. The momentum strategy in this case is similar to the S.M. strategy where the portfolio return $\tilde{r}_{t+1}$ is now given as the weighted sum:

$$\tilde{r}_{t+1} = \sum_{n=1}^{N} \tilde{y}_{n,t+1} \omega_{n,t}, \tag{2}$$

where $\tilde{y}_{n,t+1}$ is a draw from the one-period-ahead forecast distribution of the n-th industry's return $y_{n,t+1}$.[3] The weights are given as

$$\omega_{n,t} = \begin{cases} 1 & \text{if } \bar{y}_{n,t} = \max_n\{\bar{y}_{1,t}, \ldots, \bar{y}_{N,t}\} \\ -1 & \text{if } \bar{y}_{n,t} = \min_n\{\bar{y}_{1,t}, \ldots, \bar{y}_{N,t}\} \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

where $\bar{y}_{n,t}$ is the average of the fitted mean returns of the n-th industry over last 12 periods,

[3]Note that in this equation we specify a draw from the one-period ahead forecast distribution of the portfolio return. Realized returns can also be calculated alternatively using observed returns instead of $\tilde{y}_{n,t+1}$.

including time t.

To our knowledge, such a model-based momentum strategy is not considered in the literature, but it is a natural extension of the S.M. strategy. We emphasize that given our Bayesian inference procedure and given that the weights are (nonlinear) functions of the random variables $\bar{y}_{n,t}$, the underlying model and parameter uncertainty is fully taken into account.

Residual based momentum (R.M.): Next, we consider a model-based residual momentum strategy. For this portfolio, we use the fitted asset returns in the past period, invest in the assets with the highest *unexpected* returns, and go short in assets with the lowest *unexpected* returns. Unexpected returns in this strategy correspond to the model *residuals* at the investment decision time. This strategy can be seen as an extension of Blitz et al. (2011). The R.M. strategy proposed in Blitz et al. (2011) sorts the returns based on past 12 residuals from the Fama-French factor model. The assets with unexpectedly high (low) residuals are given a positive (negative) weight. The proposed R.M. strategy follows the same intuition but now for any specified model of the previous section and hence is not so restrictive as the Fama-French factor model. The R.M. weights are computed as follows:

$$\omega_{n,t} = \begin{cases} 1 & \text{if } \bar{\varepsilon}_{n,t} = \max_n\{\bar{\varepsilon}_{1,t}, \ldots, \bar{\varepsilon}_{N,t}\} \\ -1 & \text{if } \bar{\varepsilon}_{n,t} = \min_n\{\bar{\varepsilon}_{1,t}, \ldots, \bar{\varepsilon}_{N,t}\} \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

where $\bar{\varepsilon}_{n,t}$ is the average of the residuals for the n-th industry return over last 12 periods, including time t. Similar to the M.M. strategy defined above, the constructed industry industry portfolio is a weighted sum of 10 industry returns. The difference between M.M. and R.M. strategies is the use of fitted residuals indicating unexpected returns in the latter case. This can be interpreted as an error correction mechanism where portfolio weights adjust according to the deviation of the last periods' industry returns from the fitted industry return distribution.

The construction of our two strategies aims to include a plausible set of investment strategies for each model or model combination. M.M. and R.M. strategies have the advantage of providing an economic intuition of capturing estimated market trends. M.M. follows the market trends explained by the systematic component, such as common factors, and R.M. builds on return patterns that relate to the unexplained component, *i.e.*, R.M. can serve as a 'correction mechanism' when the underlying model of returns fails to represent all market dynamics.

Equally weighted portfolios: As an additional benchmark to the case of the S.M. strategy, we consider an equally weighted portfolio of combined models and strategies. We note that this portfolio differs from a model and strategy free equal weight portfolio for which one does not evaluate a measure of uncertainty. We allocate equal weight $\frac{1}{M \times S}$ to each portfolio resulting from a model and strategy pair as in (3) and (4), and we borrow at the risk-free rate in the sense that the 1-month Treasury bill rate gets weight -1. Since the portfolio weights in (3) and (4) sum up to 0, the equally weighted portfolio weights also sum up to 0. The purpose of considering this equally weighted portfolios is to identify the importance of time-variation in model and portfolio strategy performances.

Remark We have experimented with a minimum variance (M.V.) strategy, since it is widely used in applications and it is directly related to the forecasts of asset returns, volatilities and co-volatilities. However, in our empirical exercise, the results of the M.V. strategy are not explicitly included since the realized returns from this strategy were unstable for all models due to estimation uncertainty and potential ill-conditioning in variance-covariance matrix estimates, see also Michaud (1989). A fair inclusion of the results of this strategy requires more structured or 'sparse' variance-covariance matrix estimation as in Kaufmann and Schumacher (2017). This is left as a topic for further research.

4 Learning to average FDCs of models and strategies

In this section we extend the FDC approach by Billio et al. (2013) to include models as well as strategies. For more background and a survey on the evolution of the FDC approach in economics, we refer to Aastveit et al. (2018). This approach relates to the literature on dynamic prediction pools proposed in Geweke and Amisano (2010), Waggoner and Zha (2012) and Del Negro et al. (2016). However, we follow a fully Bayesian approach and make use of a different law of motion for the combination weights. The origin of all this work is the basic practice in macroeconomic and financial forecasting which consists of using a weighted combination of forecasts from many sources, say models, experts and/or large micro-data sets. In such a situation, one deals with three groups of variables: forecasts from different models, weights to combine these, and the variable of interest that is forecasted. The FDC approach gives this practice a probabilistic foundation by introducing forecast densities for different models, a weight density and a combination density. This allows for the quantification of the uncertainty of such implied density features as, in our case, mean returns, volatilities and risk of large losses.

We focus in this section on three topics. We start to discuss the specific four periods in the time-line of model estimation and portfolio construction with the implied different return variables of a portfolio strategy. We note that in a standard FDC one has forecast densities from different models that are combined to form the forecast density of the observed variable of interest (such as GDP growth, inflation) in some optimal way. In our case we deal with several constructed return variables and we discuss how and when the densities of these variables are used in the different periods of the time-line of the process. As a next step we show how the proposed FDC of model forecasts and strategy returns can be specified as a nonlinear state space model. Using very general classes of distributions in this context, the FDC model typically does not admit an analytical solution. Therefore, in order to conduct inference about this process, as a third step, we make use of numerical methods based on Bayesian sampling based filtering. Given the computational complexity of our set of models and strategies, we introduce a novel, efficient and robust filtering method, labeled the *M-Filter*. This leads to a substantial reduction in computation time using also parallel computational procedures. We refer to Appendix B for technical details.

4.1 Time-line of model estimation, construction and holding of portfolios

In Figure 2, we present a time-line of four periods that refer to estimating models for returns, constructing industry portfolios, combining models and strategies in our FDC approach and, finally, the holding of such a portfolio for a certain period and the actual return obtained from that portfolio. For convenience, we restrict our discussion to the construction of the random variables but we emphasize these are used in our simulation-based Bayesian procedure in order to construct the densities of these random variables taking full parameter and model uncertainty into account.

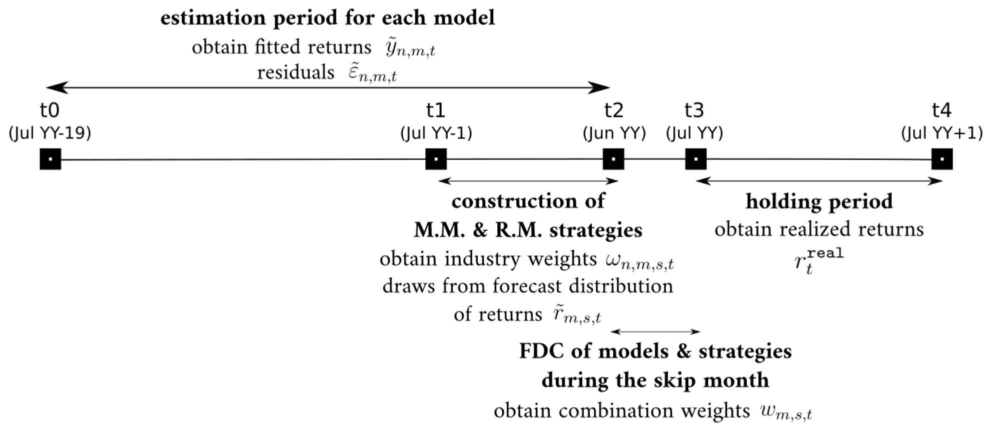
In the first two periods [t0-t2], indicated at the top of Figure 2, there are M different models estimated annually in the month of June using the preceding 240 monthly observations. The result is a set of fitted returns and obtained residuals for each industry, denoted by $\tilde{y}_{n,m,t}$ and $\tilde{\varepsilon}_{n,m,t}$, respectively, for $m = 1, \dots, M$ models. In the second period, [t1-t2] in Figure 2, S different investment strategies are formed for each model using the returns and the weights that are based on the portfolio performances in the last 12 months, including June. This strategy formation is similar to Jegadeesh and Titman (1993) and Fama and French (1993) where we construct industry weights $\omega_{n,m,s,t}$ for industry n, model m and strategy s at time t, at the end of a *skip month*, July, see also Figure 2.[4] Using equations (2)–(4), a draw from the one-period-ahead forecast distribution of the portfolio return of strategy s and model m for time $t+1$ is given by:

$$\tilde{r}_{m,s,t+1} = \sum_{n=1}^{N} \tilde{y}_{n,m,t+1} \omega_{n,m,s,t}. \tag{5}$$

We re-emphasize that our extension of the FDC approach includes an important difference compared to the standard one. In the latter case one compares the one-period-ahead forecast distribution of return, $\tilde{r}_{m,s,t+1}$, with the density of the variable of interest which is observable. In our case, we define the variable of interest, r_t, as the actual return obtained from investing one unit in the asset with maximum return and dis-investing from the asset

[4] In the literature, the skip month is often used to remove market micro-structure effects, see Asness et al. (2013). Our empirical results are robust to using the month of June for obtaining forecasts and keeping July as the skip month. The portfolio is held for 12 months starting from August every year.

Figure 2: *Time-line of model estimation, strategy construction, FDC, portfolio holding period and realized return.*



Note: 'YY' indicates the year of portfolio decision.

with minimum return. This is not observed ex ante. We define this as the full information return under the constraint that portfolio weights sum up to 0. That is, it is based on a strategy that goes long in the asset with the highest return, and goes short in the asset with the lowest return. Therefore, this full-information return can be computed as:

$$r_t = \max_n y_{n,t} - \min_n y_{n,t}. \tag{6}$$

In the third period, [t2-t3] in Figure 2, our Bayesian FDC approach approximates the distribution of (6) with the distribution of (5) (in the sense of minimizing the Kullback-Leibler divergence) in order to construct densities which are the basis for the combination approach and obtaining combination weights $w_{m,s,t}$. We explain details of this combination in the next subsections.

In the fourth period, [t3-t4] in Figure 2, we evaluate the actual returns, denoted by $r^{\texttt{real}}_{m,s,t+12}$ using alternative sets of models and strategies. In addition, we evaluate and obtain the combined realized return, $r^{\texttt{real}}_{t+12}$, over a holding period of 12 months as follows:

$$r^{\texttt{real}}_{m,s,t+12} = \sum_{t'=t+1}^{t+12} r^{\texttt{real}}_{m,s,t'} = \sum_{t'=t+1}^{t+12} \sum_{n=1}^{N} y_{n,t'} \omega_{n,m,s,t}, \tag{7}$$

$$r^{\texttt{real}}_{t+12} = \sum_{t'=t+1}^{t+12} \sum_{m=1}^{M} \sum_{s=1}^{S} r^{\texttt{real}}_{m,s,t'} w_{m,s,t} \tag{8}$$

where $y_{n,t'}$ are the realized returns for each industry, $\omega_{n,m,s,t}$ is the weight of industry n given model m and strategy s, $w_{m,s,t}$ is the weight of the combination of model m and strategy s; both types of weights are determined at time t. Realized returns in equations

(7) and (8) are then used to assess the risk-return features of all models, strategies and a combination of these.

4.2 Density combinations of model forecasts and strategy returns

In this subsection we present for the third period, [t2-t3] in Figure 2, how the FDC approach makes use of the different returns constructed from sets of models and strategies which were presented in Sections 2 and 3. The FDC model can be described as:

$$\begin{aligned} p(r_t|I) &= \int\int p(r_t, \boldsymbol{w}_t, \tilde{\boldsymbol{r}}_t|I) d\boldsymbol{w}_t d\tilde{\boldsymbol{r}}_t \\ &= \int\int p(r_t|\boldsymbol{w}_t, \tilde{\boldsymbol{r}}_t) p(\boldsymbol{w}_t) p(\tilde{\boldsymbol{r}}_t|I) d\boldsymbol{w}_t d\tilde{\boldsymbol{r}}_t, \end{aligned} \quad (9)$$

where $\boldsymbol{w}_t$ and $\tilde{\boldsymbol{r}}_t$ are the $M \times S$ matrices consisting of weights $w_{m,s,t}$ and draws from the forecast distribution $\tilde{r}_{m,s,t}$, respectively. In addition, $p(r_t|\boldsymbol{w}_t, \tilde{\boldsymbol{r}}_t)$ is specified as a combination density that explicitly incorporates the weights, $p(\boldsymbol{w}_t)$ is the weight density and $p(\tilde{\boldsymbol{r}}_t|I)$ is the joint forecast density of all M models and S strategies. Note that integrals are thus of dimension $M \times S$.

We next give content to the combination density and the weight density. Partly for convenience, we specify the combination density as a normal density[5]. This implies that there exists a model that presents the connection between the $M \times S$ forecasts from the different sources, $\tilde{r}_{m,s,t}$ with r_t as:

$$r_t = \sum_{m=1}^{M} \sum_{s=1}^{S} \tilde{r}_{m,s,t} w_{m,s,t} + \varepsilon_t, \;\; \varepsilon_t \sim NID(0, \sigma_\varepsilon^2), \;\; t = 1, \ldots, T. \quad (10)$$

The model in equation (10) contains two fundamental features: First, the matrix of weights $w_{m,s,t}$ for M models and S strategies consists of *(unobserved) random variables* so that we can model and evaluate their uncertainty. Note that one can also evaluate the correlations between the weights of the different models.

Secondly, we have added an error term ε_t which is an indication that *model incompleteness* can be modeled and evaluated. That is, as well as Bayesian learning, (10) also allows for Bayesian diagnostic analysis of misspecification. Note that with $\varepsilon_t = 0$, the density $p(r_t|\boldsymbol{w}_t, \tilde{\boldsymbol{r}}_t)$ would be a Dirac density. These two features make the proposed approach more general than Bayesian Model Averaging where the weights are posterior probabilities

[5]Different specifications of the combination density are possible. This is left as a topic of further research.

Table 1: FDC as a nonlinear state space model.

$\int\int p(r_t \mid \boldsymbol{w}_t, \tilde{\boldsymbol{r}}_t) p(\boldsymbol{w}_t) p(\tilde{\boldsymbol{r}}_t \mid I) d\boldsymbol{w}_t d\tilde{\boldsymbol{r}}_t$	
Combination density $r_t \sim NID\left(\sum_{m=1}^{M}\sum_{s=1}^{S} \tilde{r}_{m,s,t} w_{m,s,t}, \sigma_\varepsilon^2\right)$	Measurement equation $r_t = \sum_{m=1}^{M}\sum_{s=1}^{S} \tilde{r}_{m,s,t} w_{m,s,t} + \varepsilon_t,$ $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$
Link function $w_{m,s,t} = \frac{\exp(x_{m,s,t})}{\sum_{m=1}^{M}\sum_{s=1}^{S}\exp(x_{m,s,t})}$, for $m = 1, \ldots, M, \;\; s = 1, \ldots, S.$	
Markov process $x_t \sim NID\left(x_{t-1} + h(z_t), \sigma_\eta^2 I_{M\times S}\right)$	Transition equation $x_t = x_{t-1} + h(z_t) + \eta_t,$ $\eta_t \sim NID\left(0, \sigma_\eta^2 I_{M\times S}\right)$
where x_t is the $(M \times S)$-vector of $x_{m,s,t}$, $I_{M\times S}$ is the identity matrix and z_t may be included to capture (observed) economic variables believed to help explain x_t.	

that are fixed and the true model is assumed to be in the model set.

We provide the details of the weight density specification using the link function in Table 1. We use the weight density in the same way as in Billio et al. (2013), as summarized on the left hand side of Table 1. Given such a weight structure, the next point is to evaluate the model. For general forms of the weight and combination densities, it is usually not known how to evaluate these densities numerically in a direct way. In this situation one can make use of representation results, due to Billio et al. (2013), which state that the density combination model can be written as a nonlinear state space model. A summary of this result is presented for the continuous case in Table 1. This figure shows how the FDC approach is connected to filtering methods from the literature on nonlinear state space modeling and inference. We illustrate this connection in the next subsection.

Remark. We note that Casarin et al. (2018) restated the continuous case and provide a representation of the forecast density combination as a large *finite* mixture of convolutions of densities from different models. The essential step is that the combination density is now replaced by a finite mixture density. This adds flexibility to the FDC approach, we leave this as a topic for future research.

4.3 The M-Filter

Through a set of simulation studies, we show that the proposed filter is an improvement in terms of the approximation properties and computing time compared to other non-linear and non-Gaussian filters such as the Bootstrap Particle Filter (BPF) of Gordon et al.

(1993) and the Auxiliary Particle Filter (APF) of Pitt and Shephard (1999).

The combination scheme in Table 1 admits the general state space model (SSM) representation:

$$r_t \sim p(r_t|\boldsymbol{\alpha}_t), \tag{11}$$

$$\boldsymbol{\alpha}_t \sim p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}), \tag{12}$$

in which (11) and (12) describe the measurement process of the 'optimal return' r_t from equation (6) (treated as 'the dependent observation'), and the transition process of the extended state, respectively. We assume an initial state distribution $\boldsymbol{\alpha}_0 \sim p(\boldsymbol{\alpha}_0)$. Note that the extended state consists of the latent combination weights and the, potentially, fixed parameters of the system, most importantly the measure of model-strategy set incompleteness σ_ε^2. The extended state can also include appropriately specified learning parameters. We are interested in $p(\boldsymbol{\alpha}_t|r_{1:t})$, the marginal distribution of the posterior distribution of the state, called *filtering distribution* and given by

$$p(\boldsymbol{\alpha}_t|r_{1:t}) = \int p(\boldsymbol{\alpha}_{0:t}|r_{1:t}) d\boldsymbol{\alpha}_{0:t-1}. \tag{13}$$

Our novel filtering approach is summarized as follows. Firstly, the M-filter extends the particle filtering methods by not needing a resampling step. Secondly, it extends efficient importance sampling by using an on-line sequential procedure. Thirdly, in the approximation use is made of a very flexible mixture of Student's t distributions instead of the more restrictive exponential class.

We first explain our extension of the particle filter literature. These filters are based on a *recursive* formula for (13), which expresses $p(\boldsymbol{\alpha}_t|r_{1:t})$ as a function of $p(\boldsymbol{\alpha}_{t-1}|r_{1:t-1})$ and r_t, possibly time-varying. Then the computations are carried out in two steps: prediction and updating. The former step relates to the way we sample the draws at time t and the latter provides an IS correction for not using the true target density for sampling. Importantly, propagation of the particles leads to the necessity of *resampling*, as the sequential importance sampling is bound to lead to weight degeneracy problems and in consequence finally only one particle carrying the full weight. Not only might the resampling step be time consuming but it also introduces additional MC variation.[6] We avoid the propagation step by replacing it by an independent sampling step in each time period t. Here we extend the literature about importance sampling for SSM based on smoothing, e.g. EIS of Richard and Zhang (2007) and Liesenfeld and Richard (2003), or NAIS of Koopman et al. (2015). These methods are based on obtaining a good approximation to the smoothing density at each time period t and drawing from each $p(\boldsymbol{\alpha}_t|r_{1:t})$ independently. However, they are used in an off-line analysis. That is, based on a sample of a fixed size, while our primary goal is

[6] It also leads to *path degeneracy*, which is particularly problematic in the context of smoothing and in the MCMC sampling based on Particle MCMC, cf. Andrieu et al. (2010) and Lindsten et al. (2014).

on-line tracking based on filtering. We make use of independent sampling in a *sequential* way using a very *flexible* approximation density based on mixtures of Student's t densities. In order to specify our filtering method, we start with explicitly expressing (13) as follows

$$p(\boldsymbol{\alpha}_t|r_{1:t}) \propto p(r_t|\boldsymbol{\alpha}_t)p(\boldsymbol{\alpha}_t|r_{1:t-1}). \tag{14}$$

Equation (14) presents a basic Bayesian formula, where the posterior distribution of the current sate $\boldsymbol{\alpha}_t$ given all the available data $r_{1:t}$ is proportional to the prior $p(\boldsymbol{\alpha}_t)$ updated by the likelihood $p(r_t|\boldsymbol{\alpha}_t)$, where we condition upon $r_{1:t-1}$. The likelihood involves only the most recent observation r_t due to the sequential structure of the SSM. Even though we do not want to perform propagation of importance densities in the usual way of filtering procedures, we still need to keep track of the sequential structure of the SSM. We achieve this by putting a *hierarchical prior* on $\boldsymbol{\alpha}_t$, based on the empirical distribution of $\boldsymbol{\alpha}_{t-1}$.

$$p(\boldsymbol{\alpha}_t|r_{1:t}, \boldsymbol{\alpha}_{t-1}) \propto p(r_t|\boldsymbol{\alpha}_t)p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1})p(\boldsymbol{\alpha}_{t-1}). \tag{15}$$

Suppose that we have a sample $\{\boldsymbol{\alpha}_{t-1}^{(i)}\}_{i=1}^M$ from the previous time period $t-1$ so that we can approximate $p(\boldsymbol{\alpha}_{t-1})$ as $p(\boldsymbol{\alpha}_{t-1}) \approx \frac{1}{M}\sum_{i=1}^M \delta_{\boldsymbol{\alpha}_{t-1}^{(i)}}(\boldsymbol{\alpha}_{t-1})$, where $\delta_a(\cdot)$ is the Dirac measure at a. Then, (15) becomes:

$$p(\boldsymbol{\alpha}_t|r_{1:t}, \boldsymbol{\alpha}_{t-1}) \underset{\sim}{\propto} \frac{1}{M} p(r_t|\boldsymbol{\alpha}_t) \sum_{i=1}^M p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}^{(i)})\delta_{\boldsymbol{\alpha}_{t-1}^{(i)}}(\boldsymbol{\alpha}_{t-1}). \tag{16}$$

Typically we cannot draw from (16) directly and we need to resort to sampling techniques such as importance sampling (IS).

The choice of the proposal density is crucial for the performance of any Importance Sampling (IS) scheme and it has received considerable attention in the SMC literature, cf. Doucet et al. (2001), Liu (2001), Kunsch (2005) and Creal (2012). In the M-Filter we base our approximation of (16) on the *Mixture of t by Importance Sampling weighted Expectation–Maximization* (MitISEM) algorithm proposed by Hoogerheide et al. (2012) and developed in Baştürk et al. (2016). It has been shown to be able to effectively approximate complex, non-elliptical distributions thanks to two main features of this algorithm: the class of importance distributions (mixtures of multivariate Student's t distributions), and their joint optimization (with the Expectation-Maximization, EM, algorithm). The former allows to closely track distributions of nonstandard shape (multi-modal, skewed). The latter is iteratively carried out with the objective of minimizing the Kullback-Leibler (KL) divergence between the unknown true target distribution and the candidate density.

Robustness and flexibility in constructing approximations are particularly important from the filtering perspective in econometrics. For instance, stochastic volatility of many time series demonstrates itself via volatility clustering and it might be hard to efficiently capture periods of low and high volatility using standard approaches based on a single density approximation. Furthermore, especially in macro-econometrics one often observes breaks in time series which usually are very challenging to filter. We refer to the latter issue in the later part of this section.

Employing the basic MitISEM algorithm to approximate (16) means targeting the marginal posterior density $p(\boldsymbol{\alpha}_t|r_{0:1}, \boldsymbol{\alpha}_{t-1})$ with a categorical prior $\mathcal{C}(\{\boldsymbol{\alpha}_{t-1}^{(i)}\}_{i=1}^{M})$ (with equal weights). Hence, drawing from such a posterior density requires sampling the prior hyperparameters from the categorical distribution being the equally weighted sample of $\{\boldsymbol{\alpha}_{t-1}^{(i)}\}_{i=1}^{M}$. In practice, this means adopting *hierarchical Bayesian modeling*, in which at the first stage we draw $\boldsymbol{\alpha}_{t-1} \sim \mathcal{C}(\{\boldsymbol{\alpha}_{t-1}^{(i)}\}_{i=1}^{M})$, and at the second stage we draw $\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1} \sim g_t^{(H)}(\boldsymbol{\alpha}_t)$, where $g_t^{(H)}(\boldsymbol{\alpha}_t)$ is the final approximation being a mixture of H Student's t densities. The resulting sample $\{\boldsymbol{\alpha}_t^{(j)}\}_{j=1}^{N}$ becomes the empirical prior for the next time period's analysis.

Importantly, the MitISEM algorithm requires only candidate draws and IS weights, so it can simultaneously deal with several target densities. Suppose that at time t a separate target density is specified based on each draw $\boldsymbol{\alpha}_{t-1}^{(j)}$, $j = 1, \dots, M$ obtained in the previous time period, i.e.

$$p(r_t|\boldsymbol{\alpha}_t, \tilde{r}_t)p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}^{(j)}).$$

Then we construct a single approximation for these multiple targets for each time period t using MitISEM by *minimizing the average of the Kullback-Leibler (KL) divergences* between the target densities and the candidate density. In this setting the target for $\boldsymbol{\alpha}_t$ depends on $\boldsymbol{\alpha}_{t-1}^{(j)}$ but the candidate does not. We call this specific application of MitISEM for the purpose of quick filtering the M-filter algorithm. In our situation the target density of $\boldsymbol{\alpha}_t$ given $\boldsymbol{\alpha}_{t-1}$ does not crucially depend on the particular value of $\boldsymbol{\alpha}_{t-1}$, so that conditioning on the mean, variance and other characteristics of the distribution of $\boldsymbol{\alpha}_{t-1}$ suffices here. We provide the details of the algorithm in Appendix B. Note that computational efficiency gains are feasible by making use of GPU and parallel computing.

Validation and importance for typical features of economic time series: Monte Carlo experiments reported in Appendix C demonstrate a good statistical performance of the M-Filter. To illustrate its economic relevance we compare below the performance of the M-Filter and two other filters, the Bootstrap Particle Filter (BPF) of Gordon et al. (1993) and the Auxiliary Particle Filter (APF) of Pitt and Shephard (1999), on an experiment with structural breaks in the time series. We examine two cases of structural breaks in AR(1) models and we use the finite mixture scheme in Table 1 with the logistic weight specification, so that the measurement equation is nonlinear in the state process.

We simulate the following five return series with different persistence, which play the role of the draws $\tilde{r}_t$ from the forecast densities:

$$\tilde{r}_{1,t} = \frac{k}{10} + \frac{k}{10}\tilde{r}_{1,t-1} + \eta_t, \quad \eta_t \sim \mathrm{N}(0,1), \quad k = 1, \dots, 5.$$

__Table 2:__ The Mean Squared Error MSE = $\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{R}\sum_{i=1}^{R}(\tilde{\alpha}_{t,i}-\alpha_{t,i})^2\right)$, both relative to the Kalman Filter (KF). $\tilde{\alpha}_{t,i}$ denotes the posterior mean obtained in the ith replication. The results are obtained from $R = 100$ replications, with 50,000 particles for the Bootstrap Particle Filter (BPF), the Auxiliary Particle Filter (APF), and our M-Filter.

	Case 1		Case 2	
Model	MSE	Time	MSE	Time
KF	1.000	0.007	1.000	0.007
BPF	0.052	58.483	0.202	58.483
APF	0.081	68.015	0.077	68.015
M-Filter	**0.039**	**40.676**	**0.067**	**41.180**

Next, we create the measurement series r_t as a series switching between the generated series $\tilde{r}_{i,t}$, $i = 1, \ldots, 5$. We then compare the M-Filter with the BPF and APF for two different cases, varying in the number of breaks in the series, as described below. The first case has a single break/switch while the second case has two breaks/switches to emulate crisis periods.

Case 1: One switch at $t = 101$ from $\tilde{r}_1$ to $\tilde{r}_5$:

$$r_t = \begin{cases} \tilde{r}_{1,t} + \varepsilon_t & \text{for } t = 1, 2, \ldots, 100, \\ \tilde{r}_{5,t} + \varepsilon_t & \text{for } t = 101, 102, \ldots, 200, \end{cases}$$

where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ with $\sigma_\varepsilon = 0.05$.

Case 2: Two switches at $t = 101$ ($\tilde{r}_1 \to \tilde{r}_5$) and $t = 151$ ($\tilde{r}_5 \to \tilde{r}_3$):

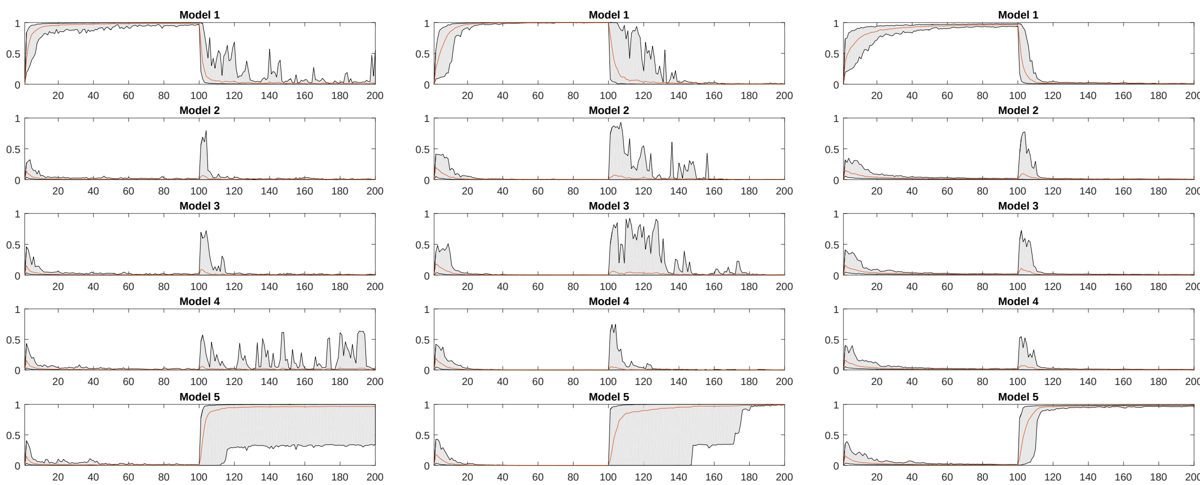
$$r_t = \begin{cases} \tilde{r}_{1,t} + \varepsilon_t & \text{for } t = 1, 2, \ldots, 100, \\ \tilde{r}_{5,t} + \varepsilon_t & \text{for } t = 101, 102, \ldots, 150, \\ \tilde{r}_{3,t} + \varepsilon_t & \text{for } t = 151, 152, \ldots, 200, \end{cases}$$

where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ with $\sigma_\varepsilon = 0.05$.

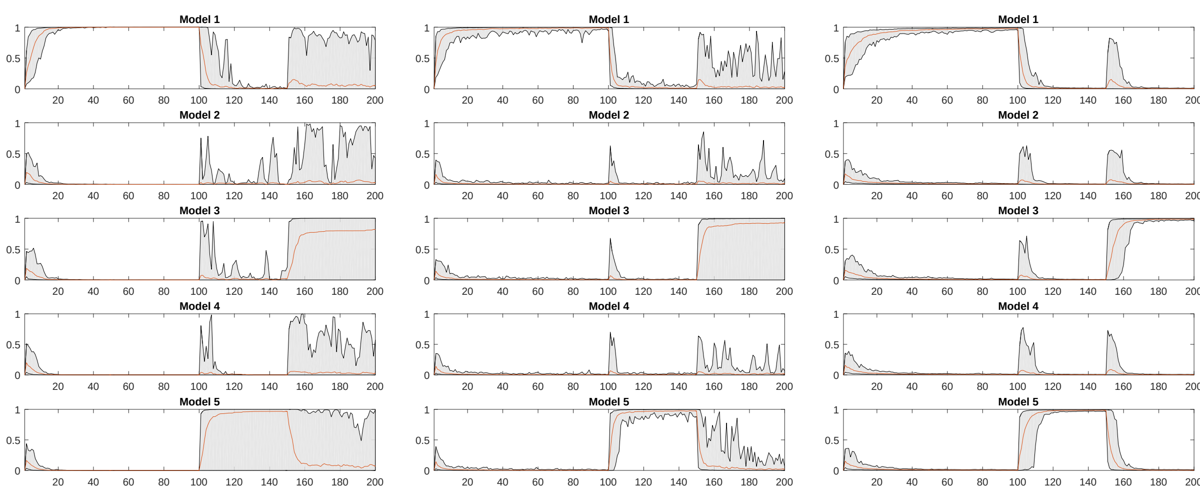
We compare the performance of the BPF, APF and M-Filter in a small MC experiment of $R = 100$ replications. Table 3(a) presents a comparison of different filters for structural breaks in AR(1) models based on the Mean Squared Error (MSE), where the error is the difference between the estimated state and the true state $r_t - \varepsilon_t$, for two different experiments. In both Case 1 and Case 2 the MSE is lowest for the M-Filter. This can be contributed to the fact that it is more precise in adapting after the shift(s), even though it requires a bit more time in adapting at the beginning of the sample. The M-Filter importance density adapts quickly at each time period after the break(s).

We next compare the weights obtained by APF and M-Filter visually. Figures 3(a)–3(c) show the model weights for Case 1. The switch in the data generating process from Model

Figure 3: *Filtered model probability weights (red lines) using the Bootstrap Particle Filter (BPF), the Auxiliary Particle Filter (APF), and our M-Filter together with the 95% credibility region (gray area) for models 1 to 5 (different rows). Top (case 1): the true model has state* $\tilde{r}_{1,t} = 0.1 + 0.1\tilde{r}_{1,t-1} + \eta_t$, $\eta_t \sim N(0,1)$ *for* $t = 1, \ldots, 100$, *and model* $\tilde{r}_{5,t} = 0.5 + 0.5\tilde{r}_{5,t-1} + \eta_t$, $\eta_t \sim N(0,1)$ *for* $t = 101, \ldots, 200$*; bottom (case 2): the true model has state* $\tilde{r}_{1,t} = 0.1 + 0.1\tilde{r}_{1,t-1} + \eta_t$, $\eta_t \sim N(0,1)$ *for* $t = 1, \ldots, 100$, *model* $\tilde{r}_{5,t} = 0.5 + 0.5\tilde{r}_{5,t-1} + \eta_t$, $\eta_t \sim N(0,1)$ *for* $t = 101, \ldots, 150$ *and model* $\tilde{r}_{3,t} = 0.3 + 0.3\tilde{r}_{3,t-1} + \eta_t$, $\eta_t \sim N(0,1)$ *for* $t = 151, \ldots, 200$.



(a) BPF weights (Case 1) (b) APF weights (Case 1) (c) M-Filter weights (Case 1)



(d) BPF weights (Case 2) (e) APF weights (Case 2) (f) M-Filter weights (Case 2)

1 to Model 5 makes it difficult for the BPF and APF to adjust quickly and one can see that the M-Filter is faster in picking up the 'break' due to the updated candidate at each time period. Figures 3(d)–3(f) illustrate Case 2, in which there are two switches in the data generating process, first from Model 1 to Model 5, and then further to Model 3. The M-Filter is the fastest in picking up the 'breaks' (particularly the second one) which again can be contributed to the updated candidate at each time period.

5 Empirical application using return data on ten US industries, 1926-2015

Our empirical analysis intends to yield valuable information on three central issues of this paper. First, does averaging of FDCs over sets of models and strategies pay off in terms of improved features of expected return and risk features? Second, does there exist useful insight from studying the dynamic patterns of the weights in these combinations, for instance, in quiet and in more volatile periods or in terms of improving the set of models and strategies? Third, what is the effect of misspecification of the model and/or strategy set on the results? More specifically, can we identify 'bad' models and strategies, and what is the effect of removing 'bad' models and strategies? Can we use diagnostic learning, economic information and/or posterior residual analysis to improve modeling and strategy choice? We note that issue two relates to learning through updating available past information while issue three deals with the robustness of our results with respect to misspecification.

Expected return and risk features using individual models and strategies: As a preliminary step we consider the performance of FDCs of three individual models: the vector autoregressive model with normal disturbances (VAR-N), the stochastic volatility model (SV), and the dynamic factor model with $K = 4$ factors and $L = 2$ lags (DFM(4,2)); all directly connected with two individual strategies (M.M. and R.M.) as discussed in section 3. We analyze the expected return and risk features of the density function of realized returns, $r_t^{\texttt{real}}$, see equation (8), using the following four indicators: expected mean return, volatility, Sharpe Ratio and largest loss during the investment period. The results are presented in Table 3 and compared with the results of a baseline S.M. strategy presented in Jegadeesh and Titman (1993), Chan et al. (1996) and Jegadeesh and Titman (2001).

The features reported in Table 3 lead to three conclusions. First, given the substantial differences between the results of the alternative model-strategy combinations for the different indicators, there does not exist a clear winning model-strategy combination in terms of all four indicators. Second, the results of the benchmark S.M. strategy are dominated by the three model-strategy combinations in several indicators with the SV model combined with R.M. outperforming the S.M. strategy in all indicators. Clearly, it pays to make use

Table 3: *Features of expected return and risk for the realized return densities using individual models and strategies.*

Model	Model Momentum (M.M.) Mean	Vol.	S.R.	L.L.	Residual Momentum (R.M.) 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**
DFM-N(4,2)	-0.05	**5.5**	-0.009	-27.4	**0.12**	**5.4**	**0.022**	-31.1

Model	Standard Momentum (S.M.) Mean	Vol.	S.R.	L.L.
–	0.09	5.7	0.016	-26.2

Note: Bold values indicate an 'equal or better' value compared to the benchmark of S.M. We report S.M. results in a single row as this strategy is not based on a model.

of a particular econometric model with a stochastic volatility component combined with an effective strategy. Third, there is one combination of a model and a strategy that clearly performs worst: the DFM-N(4,2) model in combination with the M.M. strategy is the only combination that yields a negative average return. This may be caused by a type of model misspecification that is particularly harmful for the M.M. strategy, although more research is required for the specific reasons for this very poor performance. More detailed results on the three issues are presented in the on-line Appendix. These conclusions lead naturally to our main topic of exploring the FDCs of a set of models and strategies.

5.1 Returns from FDCs using sets of models and strategies

We report the time-varying performances of several features of FDCs using sets of models and strategies in three stages. We start with the three basic model structures, VAR-N, SV and DFM-N(4,2), that constitute together the general FAVAR-SV(4,2) class. We are interested in the contribution of each component to the total results. Next, these three models are considered as a set and combined with the set of M.M. and R.M. strategies. In the second stage, we investigate whether it is better to consider a combination of two flexible models than to consider a combination of three less advanced models. We assess whether the combination of two flexible models, which are directly connected with the set of M.M. and R.M. strategies, gives improved results in terms of expected return and risk. Third, we explore whether it is effective in terms of results to choose only one model but with a very flexible parametric structure. This is the FAVAR-SV(1-4,1-2), optimized over the number of factors and lags, see below, and directly connected with the set of M.M. and R.M. strategies.

Features of expected return and risk: The four features of the empirical distributions of realized returns from different sets of models and strategies are presented in Table 4.

First, in the top panel it is shown that a FDC of three basic models and two strategies leads to improved risk features compared to individual models combined with individual strategies. The volatility and largest loss of the set of three models and two strategies are typically lower than those of the individual models. Such improvement in risk features stems from the combination of models as well as strategies. Second, the FDC of the set of three models and two strategies does not give substantially better expected mean returns than using individual models directly connected with a set of strategies. This is apparently due to the weight of the 'bad' DFM-N(4,2) model. As shown in Figure 4(a), to be discussed in more detail below, there does not exist a strong learning about the weight of the so-called 'bad' model DFM-N(4,2) in the sense that this weight remains substantial in the FDC of this set of models and strategies. We conclude that, for our data and model-strategy set, the learning mechanism for the combination weights does not effectively lower the weight of the poor performing model over time.

Given the diversity in expected return and risk results with the FDC of the set of three basic models and two strategies components, it may be a good step to obtain 'better' results by exploring a smaller set of more flexible models. As a next stage in our analysis, we explore the set of a VAR-SV and a DFM-SV model, where for the latter model we consider the case where the FDC is optimized over the number of factors from 1 to 4 and number of lags from 1 to 2. We refer to the optimized DFM-SV model as DFM-SV(1-4,1-2). The middle panel in Table 4 presents the results of the FDC of this set of models and the two strategies. Our conclusion is that the set of two flexible models and two strategies leads to better results than the set of three basic models and two strategies. Note that the results for the FDC of the set of two strategies and individual models indicate that VAR-SV good mean return features but less so for risk features while for the case of the model DFM-SV(1-4, 1-2) the opposite holds. Thus, if an investor is interested in the joint behavior of expected return and risk, then averaging over a set of flexible models and strategies is beneficial.

In the third panel we report mean return and risk features of the empirical distribution of realized returns of one model with a very flexible parametric structure, specified as a FAVAR-SV model and optimized over 4 factors and 2 lags, directly connected with the two investment strategies, leading to 16 forecast densities. Our conclusion is that choosing a set of one very flexible model and two strategies implies better mean return but also more risk than the set of two flexible models and two strategies.

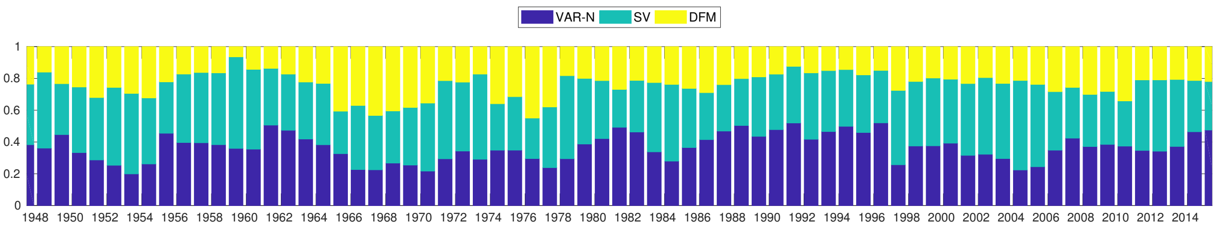
Equal weights: In the bottom panel of Table 4 the results of the equal weight scenario are shown. Equal weights perform, for our data and model-strategy set, worse than time-varying weights in terms of mean returns and Sharpe ratios. Equal weights also perform worse than S.M. in terms of mean return. We note that our FDC procedure involves a portfolio optimization that maximizes the return forecasts which is different from the equal weights scenario.

Table 4: *Features of expected return and risk for the realized return densities using sets of models and strategies.*

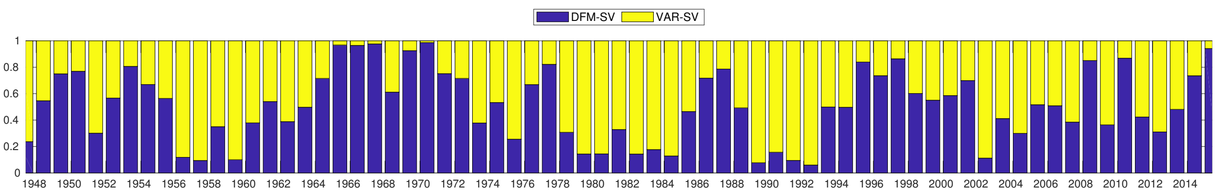
Model	Strategy	Mean	Vol.	S.R.	L.L.
Combination of three basic models and two strategies					
VAR-N & SV & DFM-N(4,2)	M.M. & R.M.	**0.10** (0.01,0.18)	**3.9** (3.6,4.2)	**0.025** (0.002,0.047)	**-23.0** (-28.8,-17.5)
Combination of strategies per model					
VAR-N	M.M. & R.M.	**0.09** (-0.03,0.20)	**4.7** (4.0.4,5)	**0.019** (-0.007,0.043)	-32.6 (-35.6,-20.9)
SV	M.M. & R.M.	**0.13** (-0.02,0.28)	**4.3** (3.9,4.6)	**0.032** (-0.005,0.065)	**-22.2** (-29.9,-16.1)
DFM-N(4,2)	M.M. & R.M.	0.03 (-0.12,0.17)	**4.3** (4.0,4.7)	0.006 (-0.028,0.041)	**-24.4** (-31.1,-16.8)
Combination of two flexible models and strategies					
VAR-SV & DFM-SV(1-4,1-2)	M.M. & R.M.	**0.15** (0.08, 0.22)	**3.7** (3.5, 3.9)	**0.041** (0.021, 0.061)	**-21.6** (-26.4, -16.4)
Combination of strategies per model					
VAR-SV	M.M. & R.M.	**0.23** (0.11, 0.35)	**4.5** (4.2, 4.9)	**0.051** (0.024, 0.080)	-37.2 (-37.3, -36.8)
DFM-SV(1-4,1-2)	M.M. & R.M.	0.06 (0.00, 0.12)	**3.4** (3.2, 3.5)	**0.018** (0.000, 0.036)	**-14.4** (-20.1, -11.0)
Combination of basic models and two strategies					
FAVAR-SV(1-4, 1-2)	M.M. & R.M.	**0.18** (0.14, 0.22)	**4.5** (4.5, 4.6)	**0.039** (0.031, 0.048)	-34.8 (-35.0, -34.6)
Benchmark models and strategies					
–	S.M.	0.09	5.7	0.016	-26.2
VAR-N, SV, DFM-N(4,2) (equal weight)	M.M. & R.M. (equal weight)	0.07 (-0.01,0.13)	**3.5** (3.3,3.8)	**0.018** (-0.002,0.038)	**-21.4** (-26.4,-16.2)
VAR-SV,DFM-SV(1-4,1-2) (equal weight)	M.M. & R.M. (equal weight)	0.07 (0.03,0.11)	**3.3** (3.2,3.4)	**0.022** (0.01,0.033)	**-13.7** (-17.8,-10.9)
FAVAR-SV(1-4, 1-2) (equal weight)	M.M. & R.M. (equal weight)	0.05 (0.02,0.07)	**3.6** (3.5,3.7)	0.013 (0.005,0.021)	**-21.6** (-24.6,-19.5)

Note: 90% credible intervals are reported in parentheses. Bold values indicate an 'equal or better' value compared to the benchmark of S.M. Equal weight denotes equally weighted models and strategies.

Figure 4: *Posterior means of model weights from FDCs using different sets of models and strategies*

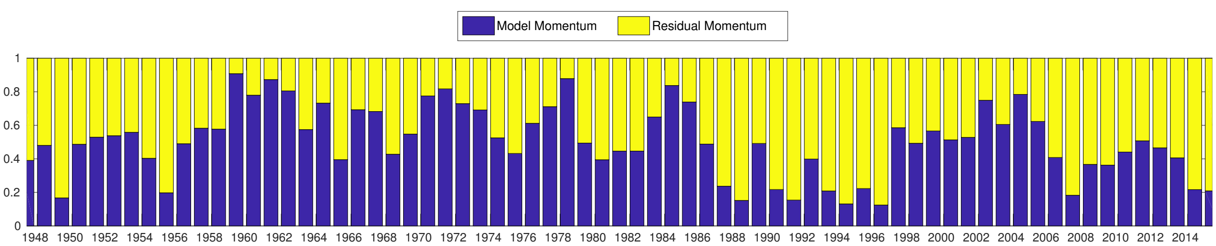


(a) Model weights of the combination of VAR-N, SV and DFM(4,2) and two strategies M.M. and R.M.

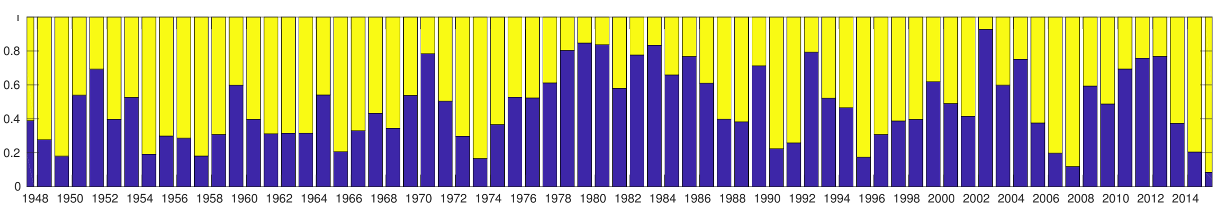


(b) Model weights of the combination of DFM-SV(1-4,1-2), VAR-SV and two strategies M.M. and R.M.

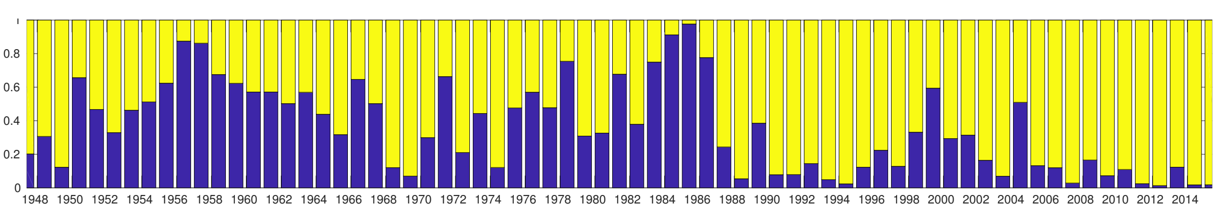
Figure 5: *Posterior means of strategy weights from FDCs using different sets of models and strategies*



(a) Strategy weights of the combination VAR-N, SV and DFM(4,2) and two strategies M.M. and R.M.



(b) Strategy weights of the combination DFM-SV(1-4,1-2), VAR-SV and two strategies M.M. and R.M.



(c) Strategy weights of the combination 8 FAVAR-SV models and two strategies M.M. and R.M.

When one compares equal weight results with those of the top, second and third panel, then it is seen that equal weights perform worse in terms of mean returns. Equal weights do lead to smaller variance and lower loss than time-varying weights for several models and also compared to S.M. The issue of diversification plays a role here. Further, the choice of the model set remains important, in both cases of equal weights and time-varying weights. Therefore, sensible a priori model selection and/or trimming of models on the basis of the results can be very beneficial. Another general conclusion is that the optimal choice of the model-strategy set depends on the preference function of the investor. In this paper we restrict ourselves to sketch results from different scenarios in terms of several features of forecast densities of portfolio returns.

Remark: Given that we obtain complete densities we also report the 90% credible intervals of the four criteria. It is seen that these intervals become smaller going from the set of three basic models to the set of two more flexible models, the set of one very flexible model and even to the equal weights case of the FAVAR-SV(1-4,1-2) model. Thus a very flexible model structure and a priori restrictions on the parameters (fixed weights) lead to more accurate estimation results. This is also relevant input information for the preference function of an investment manager.

Learning about weight dynamics and their uncertainty: It is clearly of great importance for understanding the results on expected returns and risk to explore how these features move over time. This depends on the time behavior of the FDC weights for models as well as strategies.[7] There exists time variation in the FDC weights of the two strategies as well as time variation in the FDC weights of the different sets of models. Posterior means of the model weights, presented in Figure 4, show clear time variation suggesting that autocorrelation, cross-correlation and time varying volatility data patterns are important for this weight behavior.[8] Similarly, posterior means of strategy weights, presented in Figure 5, change also substantially over time. These changes are more pronounced than those for the density combinations. A possible reason for this is that the difference between the strategies is more pronounced than the difference between the models which are all nested members of the general FAVAR-SV class of models.

We present the uncertainty in strategy weights in Figure 6 based on their 60% credibility intervals. The set of three basic models in Figure 6(a) generally leads to higher uncertainty, i.e. wider credibility intervals, than the uncertainty of two flexible models in Figure 6(b).

[7]Naturally, the choice of the particular momentum strategy or any other portfolio strategy and their implied constructed weights is also important. This is left as a topic for further research.

[8]We note that the weights per number of factors in FAVAR-SV models also change over time, as presented in Table D.9 in the on-line Appendix. 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 compared to the beginning of the sample, shown in Figure 1(b). However, a single one of the considered weights does not seem to be sufficient for these data.

Comparison of the two flexible models in Figure 6(b) and one model FAVAR-SV(1-4,1-2) in Figure 6(c) does not lead to such a clear distinction. Particularly for FAVAR-SV(1-4,1-2) in Figure 6(c), the importance of the R.M. strategies at the beginning of the recent financial crisis is confirmed with relatively low uncertainty in strategy weights.

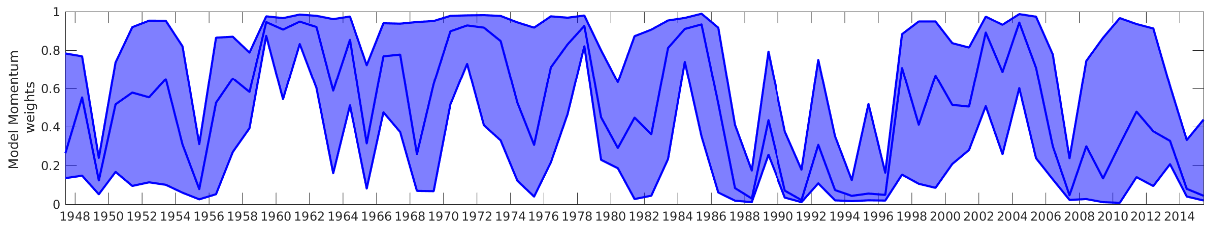
It is also interesting and relevant for policy recommendation to investigate the behavior of the R.M. weights versus the M.M. weights. In the best performing model set, that is the set consisting of the VAR-SV and DFM-SV(1-4,1-2), it is seen that the M.M. strategy remains important also in the recent crisis period. It is noteworthy that for the basic three models and the one very flexible model, the R.M. strategies are more important, particularly, around the 1990s and at the beginning of the recent financial crisis around 2008. This suggests that the R.M. strategy performs better in volatile periods. The R.M. strategy may be more robust in case of misspecification, which we leave as a topic for further research.

Our findings on the strategy weights relate to Jegadeesh and Titman (2001) and Blitz et al. (2011). Jegadeesh and Titman (2001) show that the momentum effect, indicated by the M.M. strategy, is apparent before and after the 1990s. Our results confirm this effect. Blitz et al. (2011) find that R.M. strategy is less affected by the market compared to M.M. during the financial crisis of 2008. The increased weights for R.M. in Figure 5, and the tight credibility intervals in Figure 6(c) confirm such good performance of the R.M. strategy around 2008. One explanation for this result is that the R.M. strategy is intended to take advantage of the, in absolute sense, large residuals.

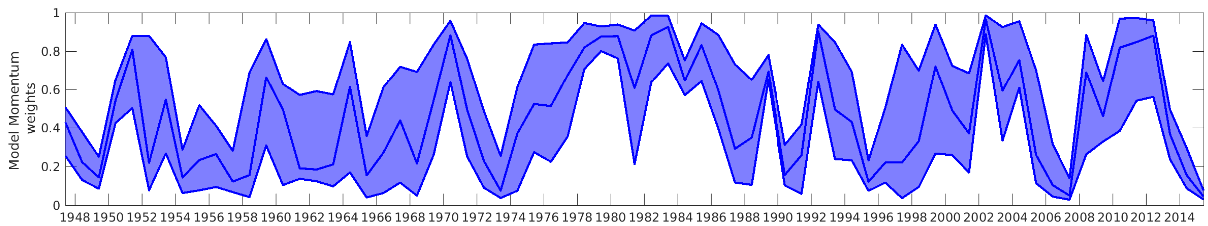
A more detailed analysis of the dynamics and learning of weights is left for further research. However, it is already seen that the dynamic patterns in model and strategy weights are very relevant pieces of information for portfolio analysis. A second important conclusion is that the learning mechanism of the method does not contain sufficient information to unambiguously give 'bad' models a very low weight when one moves through the period of observation. Diagnostic information about posterior residual behavior and poor economic performance may be useful as a complimentary source of information. This is our next topic.

Misspecification and diagnostic learning: The model and strategy sets that we considered are potentially misspecified. An important issue is how to measure this. In the statistical literature, there exist several diagnostic tests and methodologies to determine the correct number of relevant components, see McLachlan and Peel (2004, chap. 6), Frühwirth-Schnatter (2006, chap. 4) and for a recent analysis Baştürk et al. (2018). In the present paper, we follow two approaches. One is based on trimming the set of models based on their economic results, for instance, the effect that a "bad" model has on returns, like the DFM(4,2) model, as we illustrated before. This approach is related to the economics and finance literature that we discussed. A second measure follows from taking a forecasting approach. We extend the definition of σ_ε^2 from section 4.2 to forecasting and take the stan-

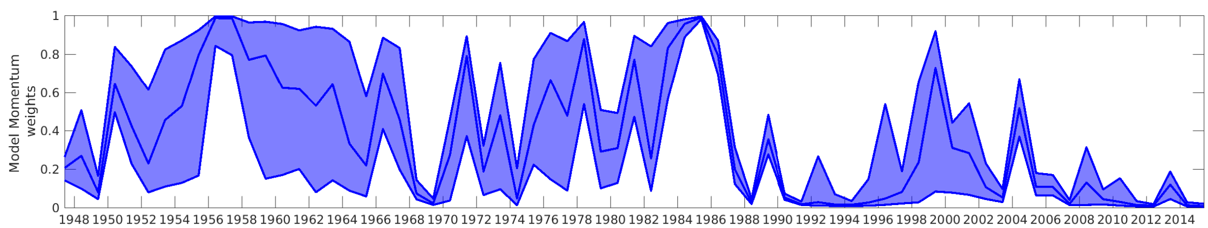
Figure 6: *Posterior means and 60% CI for FDC strategy weights using different model and strategy combinations.*



(a) Combination of VAR-N, SV and DFM(4,2) and two strategies M.M. and R.M.

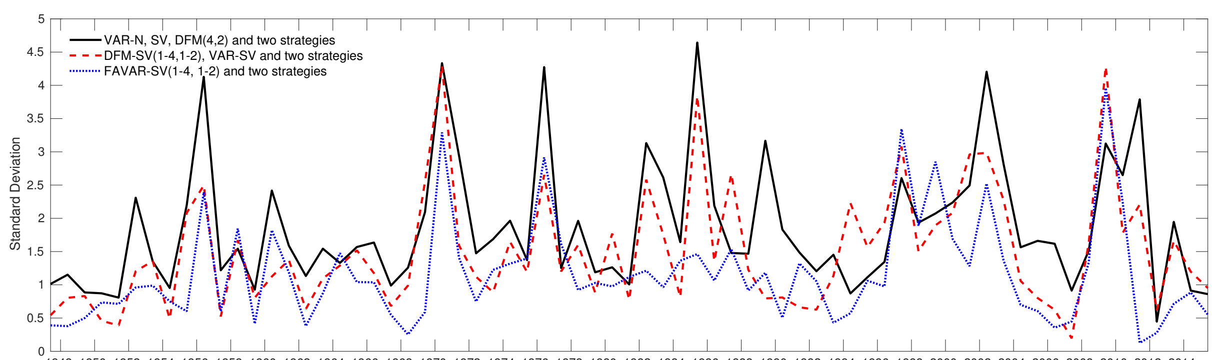


(b) Combination of DFM-SV(1-4,1-2), VAR-SV and two strategies M.M. and R.M.



(c) Combination of 8 FAVAR-SV models and two strategies M.M. and R.M.

Figure 7: *Model and strategy incompleteness measure (standard deviation of residuals).*



dard deviation of the forecast residuals as measure of incompleteness. Clearly, even when the model set is perfectly specified then this measure will be positive due to forecast errors. However, it serves a useful purpose as a relative measure comparing the performance of alternative sets of models and strategies.

Figure 7 presents the standard deviations of the forecast residuals of the FDCs in section 4.2. Standard deviations of the set of three basic models and two strategies are generally higher than those of the other two sets that we have considered. This confirms our earlier conclusion about the better fit of the flexible models compare to the basic three models. Comparison of the two flexible mixtures, VAR-SV,DFM-SV(1-4,1-2) and FAVAR-SV(1-4,1-2) does not lead to a clear result in terms of being more robust with respect to model incompleteness. It is interesting to observe that periods with high standard deviations, such as the period between 2009 and 2012, can be caused by high volatility in the data as well as by a misspecified model and strategy set. Interestingly, these periods also correspond to relatively high variations in the weights of strategies. This suggests that in case of a misspecified model-strategy set and/or a highly volatile period, it is hard to tell which model-strategy combination must be followed in order to improve features of expected return and risk. However, in periods with low standard deviations, which may be due to low volatility as well as a more complete model-strategy set, it is easier to identify a single 'winning strategy'.

6 Conclusions

Our conclusions are presented in two groups. We start with summarizing the major obtained research results. Next, we give conclusions of this paper as an advice about possible alternative investment scenarios for a professional organization in financial investment like a pension fund.

We introduced a dynamic asset-allocation approach which is specified as a forecast density combination of a set of models and momentum strategies that updates the portfolios at every decision period based on learning about their past performance. For efficient and robust Bayesian estimation of the resulting nonlinear state space model a new and efficient non-linear filter is introduced, based on the MitISEM approach, see Hoogerheide et al. (2012). It is shown that the proposed filter leads to substantial accuracy and speed gains.

Using the returns of ten US industries over an extended period, our results indicate that in volatile periods with substantial shocks it pays to make use of models that capture short-run properties through stochastic volatility components, while in quiet periods relatively less complex models receive a relevant weight. Regarding the two momentum strategies, we find that a residual momentum strategy that 'learns' from past forecast errors has higher weights compared to the simple model-based momentum strategy in volatile periods.

Thus, a time-varying equity momentum strategy leads to better performance over time. In particular, a set of models with different long and short-run dynamics together with a set of investment strategies improve two main risk properties, volatility and largest loss, of realized returns.

There are several opportunities to extend this line of research. One can consider a larger data set of industries and one can use more (economic) prior information in the model selection and in the formulation of (informative) prior distributions. This information may be particularly helpful when one intends to include mean-variance optimization methods in the analysis. Another possible extension is to assess alternative sets of combination weights of models and strategies, see also Johnstone (2012). Further, analysis of the behavior of stocks within an industry is relevant for more detailed portfolio analysis.

Advice for portfolio strategy of an investment company like a pension fund: Conditional upon our information set that consists of US industrial returns between 1926M7 and 2015M6 and the proposed set of dynamic models and data driven portfolio strategies, the properties of mean return and risk of a portfolio improve using our FDCs with sets of models and strategies. Using single models and strategies, including the standard momentum strategy, yields lower mean returns and more risk in most cases. Finally, learning weights of the FDCs of sets of models and strategies should be carefully incorporated in exploring alternative scenarios of portfolio strategies. The proposed time-varying weights of the set of two flexible models and two strategies outperforms the equally weighted combination of models and strategies in the sense of better return and risk trade-off. Compared to fixed weights, these gains are rather pronounced in volatile periods.

We end this paper with the statement that it, of course, depends on the return-risk trade-off preference of a trader how to deal with these recommendations. We do re-emphasize, however, that our proposed data-driven FDC approach for combining models and strategies does not require a fully specified utility function that is particular for a trader.

References

Aastveit K A, Mitchell J, Ravazzolo F, Van Dijk H K. 2018. The Evolution of Forecast Density Combinations in Economics. Technical Report 18-069/III, Tinbergen Institute. *Oxford Research Encyclopedia in Economics and Finance* Forthcoming.

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

Andrieu C, Doucet A, Holenstein R. 2010. Particle Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72**: 269–342.

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

Baştürk N, Grassi S, Hoogerheide L, Opschoor A, Van Dijk H K. 2017. The R package MitISEM: Efficient and robust simulation procedures for Bayesian inference. *Journal of Statistical Software, Articles* **79**: 1–40. ISSN 1548-7660.

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, Van Dijk H K. 2016. Parallelization Experience with Four Canonical Econometric Models Using ParMitISEM. *Econometrics* **4**: 1–11.

Baştürk N, Grassi S, Hoogerheide L, Van Dijk H K. 2016. Time-varying Combinations of Bayesian Dynamic Models and Equity Momentum Strategies. Technical Report 16-099/III, Tinbergen Institute.

Baştürk N, Hoogerheide L F, Van Dijk H K. 2018. Bayes Estimates of Multimodal Density Features in DNA and Economic Data. Technical report, Tinbergen Institute.

Bernanke S B, Boivin J, Eliasz P. 2005. Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach. *The Quarterly Journal of Economics* **120**: 387–422.

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

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

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

Carvalho C M, Lopes H F, Aguilar O. 2011. Dynamic Stock Selection Strategies: A Structured Factor Model Framework. In Bernardo J, Bayarri M, Berger J, Dawid A, Heckerman D, Smith A, West M (eds.) *Bayesian Statistics 9*. Oxford: Oxford University Press, 1–21.

Casarin R, Grassi S, Ravazzolo F, van Dijk H K. 2018. Predictive Density Combinations with Dynamic Learning for Large Data Sets in Economics and Finance. Discussion Paper, Tinbergen Institute.

Chan J, Leon-Gonzalez R, Strachan R W. 2018. Invariant Inference and Efficient Computation in the Static Factor Model. *Journal of the American Statistical Association* **113**: 819–828.

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.

Del Negro M, Hasegawa R B, Schorfheide F. 2016. Dynamic Prediction Pools: An Investigation of Financial Frictions and Forecasting Performance. *Journal of Econometrics* **192**: 391–405.

DeMiguel V, Garlappi L, Uppal R. 2007. Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *The Review of Financial Studies* **22**: 1915–1953.

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

Engle R, Colacito R. 2006. Testing and Valuing Dynamic Correlations for Asset Allocation. *Journal of Business & Economic Statistics* **24**: 238–253.

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. 2015. A Five-Factor Asset Pricing Model. *Journal of Financial Economics* **116**: 1–22.

Frühwirth-Schnatter S. 2006. *Finite Mixture and Markov Switching Models*. Springer Verlag.

Frühwirth-Schnatter S, Lopes H F. 2018. Sparse Bayesian Factor Analysis when the Number of Factors is Unknown. ArXiv preprint 1804.04231.

Garlappi L, Uppal R, Wang T. 2006. Portfolio Selection with Parameter and Model Uncertainty: A Multi-prior Approach. *The Review of Financial Studies* **20**: 41–81.

Geweke J, Amisano G. 2010. Optimal Prediction Pools. *Journal of Econometrics* **164**: 130–141.

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

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

Gruber L F, West M. 2017. Bayesian Online Variable Selection and Scalable Multivariate Volatility Forecasting in Simultaneous Graphical Dynamic Linear Models. *Econometrics and Statistics* **3**: 3–22.

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.

Johnstone D J. 2012. Log-optimal Economic Evaluation of Probability Forecasts. *Journal of the Royal Statistical Society: Series A* **175**: 661–689.

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.

Kastner G, Frühwirth-Schnatter S, Lopes H F. 2017. Efficient Bayesian Inference for Multivariate Factor Stochastic Volatility Models. *Journal of Computational and Graphical Statistics* **26**: 905–917.

Kaufmann S, Schumacher C. 2017. Identifying Relevant and Irrelevant Variables in Sparse Factor Models. *Journal of Applied Econometrics* **32**: 1123–1144.

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

Koopman S J, Lucas A, Scharth M. 2015. Numerically Accelerated Importance Sampling for Nonlinear Non-Gaussian State Space Models. *Journal of Business and Economic Statistics* **33**: 114–127.

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.

Lindsten F, Jordan M, Schön T B. 2014. Particle Gibbs with Ancestor Sampling. *Journal of Machine Learning Research* **15**: 2145–2184.

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.

McLachlan G, Peel D. 2004. *Finite Mixture Models.* John Wiley & Sons.

Michaud R O. 1989. The Markowitz Optimization Enigma: Is 'Optimized' Optimal? *Financial Analysts Journal* **45**: 31–42.

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* **31**: 1312–1332.

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

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

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

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

Talih M, Hengartner N. 2005. Structural Learning with Time-varying Components: Tracking the Cross-section of Financial Time Series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**: 321–341.

Waggoner D F, Zha T. 2012. Confronting Model Misspecification in Macroeconomics. *Journal of Econometrics* **171**: 167–184.

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

Zhou X, Nakajima J, West M. 2014. Bayesian Forecasting and Portfolio Decisions Using Dynamic Dependent Sparse Factor Models. *International Journal of Forecasting* **30**: 963–980. ISSN 0169-2070.

Online Appendix

Appendix A Models within the FAVAR-SV class

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

A.1 Linear and Gaussian Dynamic Factor Model (DFM)

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

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

To estimate this model we assume the following priors.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\overline{\mathbf{\Phi}} \sim \mathrm{N}(\mathrm{vec}(\tilde{\Phi}), \mathrm{Q} \otimes \tilde{\Omega}), \tag{A.18}$$

   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.

A.2 Linear Dynamic Factor Model with Stochastic Volatility (DFM-SV)

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

$$\begin{aligned} \mathbf{y}_t &= \Lambda \mathbf{f}_t + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t &\sim \mathrm{N}(0, \Sigma_t), \\ \mathbf{f}_t &= \phi_1 \mathbf{f}_{t-1} + \ldots + \phi_L \mathbf{f}_{t-L} + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim \mathrm{N}(0, \mathrm{Q}), \end{aligned} \tag{A.19}$$

and specifying a time-varying variance-covariance matrix:

$$\Sigma_t = \begin{pmatrix} \sigma^2_{11,t} & 0 & \dots & 0 \\ 0 & \sigma^2_{22,t} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2_{NN,t} \end{pmatrix}, i = 1, \dots, \text{N}. \tag{A.20}$$

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

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

Starting from equation (A.19) and rearranging, we get $\boldsymbol{\varepsilon}_t = \mathbf{y}_t - \Lambda \mathbf{f}_t = \mathbf{y}^*_t$. Taking the squares plus an offset constant we obtain

$$\begin{aligned} \mathbf{y}^{**}_t &= \log[(\mathbf{y}^*_t)^2 + \bar{c}], \\ \mathbf{y}^{**}_t &= 2\mathbf{h}_t + \boldsymbol{e}_t, \\ \mathbf{h}_t &= \boldsymbol{\mu} + \boldsymbol{\psi}\mathbf{h}_{t-1} + \xi_t, \quad \xi_t \sim \text{N}(0, \Gamma), \end{aligned} \tag{A.21}$$

where $\boldsymbol{e}_t = \log(\boldsymbol{\varepsilon}_t)$ follows the $\chi^2(1)$ distribution. Therefore, the standard Kalman filter and smoother cannot be adopted, cf. Carter and Kohn (1994). To solve this problem Kim et al. (1998) employ a data augmentation approach and introduce a new state variable $\mathbf{s}_T = \{\mathbf{s}_1, \dots, \mathbf{s}_t\}$, so that the linear, non-Gaussian state space model (A.21) can be rewritten as conditionally linear Gaussian. Then, the distribution of $\boldsymbol{e}_t$ can be approximated as

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

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

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

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

$$\Pr(\mathbf{s}_t = j | \mathbf{y}^{**}_t, \mathbf{h}_t) \propto q_j f_{\text{N}}(\mathbf{y}^{**}_t | 2\mathbf{h}_t + \tau_j - 1.2704, \nu^2_j), \tag{A.22}$$

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

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

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

while for γ_{ii}^{-2} we put

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

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

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

1) Initialize $\mathbf{f}_t^{(0)}$, $\mathbf{h}_t^{(0)}$, $\Lambda_t^{(0)}, \Sigma^{(0)}, \mathrm{Q}^{(0)}$.

2) Draw latent factors $\mathbf{f}_t$ from $p(\mathbf{f}_t|, \Lambda, \mathrm{Q}, \Sigma_t, \mathbf{h}_t)$ using the FF-BS algorithm described in Carter and Kohn (1994).

3) Conditionally on $\mathbf{h}_t$ and Λ, draw the indicator variable $\mathbf{s}_t$ for the mixture according to (A.22).

4) Draw a sequence of stochastic volatilities $\mathbf{h}_t$, $t = 1, \ldots, T$ from $p(\mathbf{h}_t|\Lambda, \mathbf{f}_t, \mathbf{s}_t, \boldsymbol{\mu}, \boldsymbol{\psi})$ from the conditional linear and Gaussian system using the method of Carter and Kohn (1994).

5) Draw the stochastic volatility variances γ_{ii}^2 from $p(\gamma_{ii}^2|h_{it}, \mu_i, \psi_i)$ from the following posterior:

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

6) Draw the SV parameters jointly:

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

where

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

and

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

7) Go to step 2.

A.3 Linear Dynamic Factor Model with Two Stochastic Volatility Components (DFM-SV2)

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

$$\begin{array}{ll}\mathbf{y}_t = \Lambda\mathbf{f}_t + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t \sim \mathrm{N}(0, \Sigma_t),\\ \mathbf{f}_t = \phi_1\mathbf{f}_{t-1} + \ldots + \phi_L\mathbf{f}_{t-L} + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t \sim \mathrm{N}(0, \mathrm{Q}_t),\end{array} \tag{A.23}$$

with the idiosyncratic errors defined as in equation (A.20) and latent error variances is given by:

$$\mathrm{Q}_t = \begin{pmatrix}\eta_{11,t}^2 & 0 & \ldots & 0\\ 0 & \eta_{22,t}^2 & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & \eta_{KK,t}^2\end{pmatrix}, i = 1, \ldots, K, \tag{A.24}$$

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

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

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

A.4 Factor Augmented VAR models with Stochastic Volatility Components (FAVAR-SV2)

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

$$\begin{array}{ll}\mathbf{y}_t = \beta\mathbf{x}_t + \Lambda\mathbf{f}_t + \boldsymbol{\varepsilon}_t & \boldsymbol{\varepsilon}_t \sim \mathrm{N}(0, \Sigma_t),\\ \mathbf{f}_t = \phi_1\mathbf{f}_{t-1} + \ldots + \phi_L\mathbf{f}_{t-L} + \boldsymbol{\eta}_t & \boldsymbol{\eta}_t \sim \mathrm{N}(0, \mathrm{Q}_t).\end{array} \tag{A.25}$$

The FAVAR model extends the state equation by defining $\mathbf{x}_t$ as a vector of the lagged dependent variables. This leads to a VAR form in the state equation of (A.25):

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

see also Stock and Watson (2005). Conditional on the latent states, the estimation of the VAR parameters $\boldsymbol{\beta}$ is similar to that of the univariate linear regression model, hence Bayesian inference is standard. The two proposed FAVAR models are defined by a stochastic volatility component in the idiosyncratic disturbances (FAVAR-SV) and a stochastic volatility components in the idiosyncratic and latent disturbances (FAVAR-SV2). Note that the FAVAR-SV (and FAVAR-SV2) model simplifies to a DFM model in Section A.1 when $\beta = 0$, a VAR model if factor coefficient $\Lambda = 0$, and a stochastic volatility model when both $\beta = 0$ and $\Lambda = 0$. Hence DFM, VAR and SV models listed together constitute parts of the FAVAR-SV (and FAVAR-SV2) models. We refer to the earlier sections of this appendix for the inference on the SV components conditionally on the remaining parameters.

Appendix B The M-Filter

Below we present the details of the recursion for the proposed M-Filter from Section 4. In the description we treat the estimated parameter vector θ as known and we omit it for the sake of notation. For a detailed discussion of the general MitISEM procedure we refer to Hoogerheide et al. (2012).

Algorithm Below we present the proposed M-Filter in algorithmic form.

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

2) **Recursion**. For $t = 1, \ldots, T$ construct the candidate density $g_t(\boldsymbol{\alpha}_t)$ using the MitISEM algorithm as follows.

   a) **Initialization**: Simulate draws $\boldsymbol{\alpha}_t^{(j)}$, $j = 1, \ldots, M$, from a 'naive' candidate distribution with density $g_t^{(0)}(\cdot)$ (e.g. a Student's t distribution with $v = 5$ degrees of freedom).

   Compute the corresponding IS weights:

$$\tilde{w}_t^{(j)} = \frac{p(r_t|\boldsymbol{\alpha}_t^{(j)})p(\boldsymbol{\alpha}_t^{(j)}|\boldsymbol{\alpha}_{t-1}^{(j)})}{g_t^{(0)}(\boldsymbol{\alpha}_t^{(j)})},$$

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

   b) **Adaptation**: Use the draws $\boldsymbol{\alpha}_t^{(j)}$ and the weights $\tilde{w}_t^{(j)}$ from the naive distribution $g_t^{(0)}(\cdot)$ to IS estimate the mean and covariance matrix of the target distribution. Use these estimates as the mode and the scale matrix of the Student's t adapted density $g_t^{(a)}(\cdot)$. Draw a sample $\boldsymbol{\alpha}_t^{(j)}$ from $g_t^{(t)}(\cdot)$ and compute the IS weights for this sample.

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

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

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

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

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

d.4) Draw $\boldsymbol{\alpha}_t^{(j)}$ from the new mixture $g_t^{(h+1)}(\cdot)$ and evaluate the corresponding normalized importance weights $w_t^{(j)}$, $j = 1, \dots, M$.

d.5) Calculate the coefficient of variation $CV^{(h+1)}$ of the normalized weights $w_t^{(j)}$, $j = 1, \dots, M$.

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

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

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

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

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

Appendix C Simulation results for M-filter

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

C.1 Local level model

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

$$\begin{aligned} y_t &= \alpha_t + \varepsilon_t \qquad \varepsilon_t \sim \mathrm{N}(0, \sigma_\varepsilon^2), \\ \alpha_t &= \alpha_{t-1} + \eta_t \quad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2). \end{aligned} \tag{C.26}$$

that is a linear and Gaussian model often used 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 experiments, we fix the latent state variance as $\sigma_\eta^2 = 0.1$ and we define four different levels for the state variance σ_ε^2, corresponding to four levels of the Noise to Signal Ratio (NtS): 0.1, 0.5, 1 and 2.5. We note that the exact likelihood of the model in equation (C.26) can be calculated using the KF, and we can compare the exact likelihood of this model with the remaining non-linear filters. This allows to assess the degree of the likelihood bias in the non-linear filters, including the proposed M-Filter.

Table C.5 reports the results for the model in equation (C.26). 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 BPF 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 BPF and APF.

C.2 Stochastic volatility model

The second model is the stochastic volatility (SV) model (Kim et al., 1998) given by:

$$\begin{aligned} y_t &= e^{(\alpha_t/2)}\varepsilon_t \qquad\qquad \varepsilon_t \sim \mathrm{N}(0, \sigma_\varepsilon^2), \\ \alpha_t &= \mu + \phi\alpha_{t-1} + \eta_t \quad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2), \end{aligned} \tag{C.27}$$

Table C.5: *Monte Carlo results for R =100 replications of the linear and Gaussian model of equation* (C.26) *with T=100. Kalman Filter (KF), Bootstrap Particle Filter (BPF), Auxiliary Particle Filter (APF) and MitISEM Filter (M-Filter) with 50,000 particles. The table reports Log-Likelihood Bias (LLB) with respect to KF. Absolute deviation defined as* $Bias = 1/T\sum_{t=1}^{T}\left(1/R\sum_{i=1}^{R}|\tilde{\alpha}_{t,i}-\alpha_{t,i}|\right)$ *relative to the KF. The table also reports the variability defined as* $Var = 1/T\sum_{t=1}^{T}\left(1/R\sum_{i=1}^{R}(\tilde{\alpha}_{t,i}-\alpha_{t,i})^2\right)$ *relative to the KF. The final column reports the computing time in seconds for the four filters.*

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

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, that corresponds to different coefficient of variation (CV) of the volatility $h = \sigma_\eta^2 \exp(\alpha_t)$ defined as:

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

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

Table C.6 reports the results for the SV model of equation (C.27) with $\phi = 0.98$ and different values of σ_η^2 that corresponds to CV $= 0.1, 0.5, 1, 2$. In all cases the KF is the worst filter due to being a linear and Gaussian filter. The M-Filter performs similarly to the BPF 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 BPF, APF and M-Filter.

Table C.6: *Monte Carlo results for R=100 replications of the SV model in equation* (C.27) *with T=100,* $\phi = 0.98$ *and* $CV = 0.1, 0.5, 1, 2.5$. *The Kalman Filter (KF), Bootstrap Particle Filter (BPF), Auxiliary Particle Filter (APF) and MitISEM Filter (M-Filter) with 50,000 particles. The table reports the absolute deviation Bias* $= 1/T\sum_{t=1}^{T}\left(1/R\sum_{i=1}^{R}|\tilde{\alpha}_{t,i}-\alpha_{t,i}|\right)$ *as a ratio to the KF. The table also report the variability* $Var = 1/T\sum_{t=1}^{T}\left(1/R\sum_{i=1}^{R}(\tilde{\alpha}_{t,i}-\alpha_{t,i})^2\right)$ *as a ratio to the KF. The final column reports the computing time in seconds for the four filters for different CV.*

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

C.3 Dynamic Factor Model

The last model we examine is a Dynamic Factor Model (DFM) given by:

$$\begin{aligned} \mathbf{y}_t &= \Lambda \mathbf{f}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathrm{N}(0, \Sigma), \\ \mathbf{f}_t &= \Phi_1 \mathbf{f}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathrm{N}(0, \mathrm{Q}). \end{aligned} \tag{C.28}$$

where $\mathbf{y}_t$ is a $N \times 1$ vector of time series,the $K \times 1$ vector $\mathbf{f}_t$ contains unobservable factors with one lag where Φ_1 is a $K \times K$ matrix of autoregressive coefficients, Λ is an $N \times K$ matrix of factor loadings. Finally, $\boldsymbol{\varepsilon}_t$ is an $N \times 1$ i.i.d. vector of idiosyncratic disturbances and $\boldsymbol{\eta}_t$ is an $K \times 1$ i.i.d. vector of latent disturbances.

The model in equation (C.28) is linear and Gaussian and as in equation (C.26) the KF is the optimal filter. As before we compare the performance of the non-linear filters against the KF for different number of factors.

Table C.7 reports the results for the model in equation (C.28) for $R = 100$ Monte Carlo replication, $N = 20$ series and $K = 2, 4, 6, 10$ factors. In all simulations experiments the following simulation setting is used: Λ is a $N \times K$ matrix with zeros on the $K \times (K-1)/2$ upper-diagonal elements and the remaining elements of Λ are simulated from independent standard normal distributions; Φ_1 is a diagonal matrix with elements simulated from independent uniform distributions between [0,1); Σ is a diagonal matrix with elements simulated from a uniform distribution between [0,2.5]; Q is a diagonal matrix for which the simulations are performed in two steps. First, we simulate $\tilde{\mathrm{Q}} = \Psi\Psi'$ where Ψ is a $K \times K$ upper triangular matrix with elements simulated from independent uniform distributions in [0,1]. The diagonal elements of Q are defined as the diagonal elements of $0.1 \times \tilde{\mathrm{Q}}^{-1}$.[9]

Due to the linear and Gaussian model structure in equation (C.28), KF leads to the best results in terms of the speed and accuracy, but the non-linear filter are in line with the KF. The M-Filter performs better then both BPF and APF, with substantially lower bias and variance. The M-Filter has also the lowest likelihood bias compared to the other non-linear and non-Gaussian filters. In terms of the computing time it increases with the number of factors for all the filters. In all cases, however, M-Filter requires less computing time compared than the BPF and APF.

Table C.7: *Monte Carlo results for R=100 replications of the DFM with T=100, N = 20 and $K = 2, 4, 6, 10$ latent factors. Kalman Filter (KF), Bootstrap Particle Filter (BPF), Auxiliary Particle Filter (APF) and MitISEM Filter (M-Filter) with 50,000 particles. The table reports Log-Likelihood Bias (LLB) with respect to KF. Absolute deviation defined as $Bias = 1/T\sum_{t=1}^{T}\left(1/R\sum_{i=1}^{R}|\tilde{\alpha}_{t,i} - \alpha_{t,i}|\right)$ relative to the KF. The table also reports the variability defined as $Var = 1/T\sum_{t=1}^{T}\left(1/R\sum_{i=1}^{R}(\tilde{\alpha}_{t,i} - \alpha_{t,i})^2\right)$ relative to the KF. The final column reports the computing time in seconds for the four filters and for $K = 2, 4, 6, 10$.*

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

[9] Our general conclusions hold under different parameter values such as $\Phi_1 = 0.9$ and different specifications for the non-zero elements of Λ.

Appendix D Additional empirical results

D.1 Return and risk features from model-strategy pairs: additional results

We first summarize the realized return and risk properties of several model and investment strategies together with those of the S.M. strategy. We consider eight sets of models reported in Table D.8.[10]

Table D.8: *List of combined models sorted by increasing complexity.*

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

For each DFM model and FAVAR 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 M.M. and R.M.. In total, we estimate 40 combinations of DFM and FAVAR models, 2 VAR models (VAR and VAR-SV), and a 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 M.M. strategy and a R.M. strategy. Hence we obtain 86 combinations of model and investment strategy specifications. Realized returns from these 86 combinations are summarized in Appendix D, Table D.9 in detail. We present a selection of these results in Table D.10.

Mean realized returns differ substantially over alternative model and strategy specifications, as shown in Table D.10. In terms of these mean realized returns, M.M. strategy gives poor results for simple VAR-N and DFM-N models. The complex model structures, DFM-SV2 and FAVAR-SV2 do not necessarily lead to higher mean returns compared to more basic models DFM-SV and FAVAR-SV for both strategies. That is, the SV2 component in factor residuals 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

[10]For all models, Bayesian inference is performed with 5000 burn-in and 5000 posterior draws.

Table D.9: *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 M.M. and R.M.. S.M. 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 S.M.. K is the number of factors and L is the number of lags.*

		Model Momentum				Residual Momentum			
	(K, L)	Mean	Vol.	S.R.	L.L.	Mean	Vol.	S.R.	L.L.
VAR-N	–	0.02	**5.0**	0.005	**-24.1**	**0.09**	5.8	0.015	-35.0
SV	–	**0.10**	**5.1**	0.019	-34.7	**0.11**	**5.6**	0.019	**-26.0**
VAR-SV	–	**0.12**	**4.5**	**0.028**	-20.2	**0.13**	5.8	**0.021**	-37.4
DFM-N	(1,1)	-0.04	**4.9**	-0.009	**-20.0**	**0.13**	**5.7**	**0.023**	-34.4
	(1,2)	-0.04	**4.9**	-0.009	**-20.0**	**0.13**	**5.7**	**0.022**	-34.4
	(2,1)	-0.13	**5.2**	-0.024	**-25.4**	**0.10**	**5.6**	0.017	-34.0
	(2,2)	-0.11	**5.2**	-0.020	**-24.2**	**0.10**	**5.6**	0.017	-34.1
	(3,1)	-0.14	**5.4**	-0.027	**-23.7**	**0.09**	**5.5**	0.017	-33.7
	(3,2)	-0.08	**5.4**	-0.016	**-23.3**	0.08	**5.4**	0.015	-33.1
	(4,1)	-0.07	**5.5**	-0.013	-26.7	**0.10**	**5.4**	0.018	-31.3
	(4,2)	-0.05	**5.5**	-0.009	-27.4	**0.12**	**5.4**	**0.022**	-31.1
DFM-SV	(1,1)	0.04	**5.0**	**0.007**	**-20.0**	**0.11**	5.8	0.019	-37.1
	(1,2)	0.04	**5.0**	**0.008**	**-20.0**	**0.10**	5.8	0.018	-37.1
	(2,1)	-0.04	**5.2**	-0.009	**-22.0**	**0.15**	**5.7**	**0.026**	-36.3
	(2,2)	-0.05	**5.2**	-0.009	**-22.0**	**0.15**	**5.7**	**0.027**	-36.6
	(3,1)	0.00	**5.2**	0.000	**-21.2**	**0.14**	**5.4**	**0.026**	-33.0
	(3,2)	0.03	**5.2**	0.005	**-20.8**	**0.16**	**5.4**	**0.030**	-32.8
	(4,1)	**0.12**	**5.4**	**0.023**	**-20.8**	0.05	**5.4**	0.009	-31.8
	(4,2)	**0.12**	**5.4**	**0.023**	**-21.7**	0.06	**5.4**	0.011	-31.1
DFM-SV2	(1,1)	0.07	**4.6**	0.014	**-18.2**	0.06	**5.5**	0.010	-37.4
	(1,2)	0.07	**4.6**	0.014	**-18.2**	0.06	**5.5**	0.010	-37.4
	(2,1)	-0.01	**4.8**	-0.002	**-22.8**	0.08	**5.5**	0.015	-37.4
	(2,2)	-0.02	**4.8**	-0.003	**-22.8**	**0.09**	**5.5**	0.016	-37.4
	(3,1)	0.02	**5.0**	0.005	-27.1	-0.02	**5.5**	-0.003	-37.4
	(3,2)	0.03	**5.0**	0.006	-27.1	-0.02	**5.5**	-0.003	-37.4
	(4,1)	0.07	**5.7**	0.013	-32.3	0.00	**5.2**	0.000	-37.4
	(4,2)	0.07	**5.7**	0.013	-32.3	0.00	**5.2**	0.000	-37.4
FAVAR-SV	(1,1)	0.08	**4.6**	0.018	**-18.3**	0.06	**5.5**	0.011	-37.4
	(1,2)	0.08	**4.6**	0.018	**-18.3**	0.06	**5.5**	0.011	-37.4
	(2,1)	-0.03	**4.9**	-0.005	**-23.1**	0.08	**5.5**	0.015	-37.4
	(2,2)	-0.03	**4.9**	-0.006	**-23.5**	**0.09**	**5.5**	0.016	-37.4
	(3,1)	**0.09**	**5.0**	0.018	**-25.3**	-0.02	**5.5**	-0.005	-37.4
	(3,2)	0.08	**5.0**	0.017	**-25.7**	-0.02	**5.5**	-0.004	-37.4
	(4,1)	0.08	**5.7**	0.014	-32.3	0.03	**5.2**	0.005	-37.4
	(4,2)	0.08	**5.7**	0.015	-32.3	0.02	**5.2**	0.005	-37.4
FAVAR-SV2	(1,1)	**0.09**	**4.6**	0.019	**-18.3**	0.06	**5.5**	0.011	-37.4
	(1,2)	0.08	**4.6**	0.018	**-18.3**	0.06	**5.5**	0.011	-37.4
	(2,1)	-0.03	**4.9**	-0.005	**-23.5**	**0.09**	**5.5**	0.016	-37.4
	(2,2)	-0.03	**4.9**	-0.005	**-23.8**	0.08	**5.5**	0.015	-37.4
	(3,1)	0.08	**5.0**	0.017	**-25.6**	-0.03	**5.5**	-0.005	-37.4
	(3,2)	0.08	**5.0**	0.017	**-25.3**	-0.02	**5.5**	-0.004	-37.4
	(4,1)	0.08	**5.7**	0.014	-32.3	0.03	**5.2**	0.005	-37.4
	(4,2)	0.08	**5.7**	0.014	-32.3	0.03	**5.2**	0.005	-37.4

Table D.10: *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 different models and strategies in Section 2. S.M. 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 S.M.. K is the number of factors and L is the number of lags.*

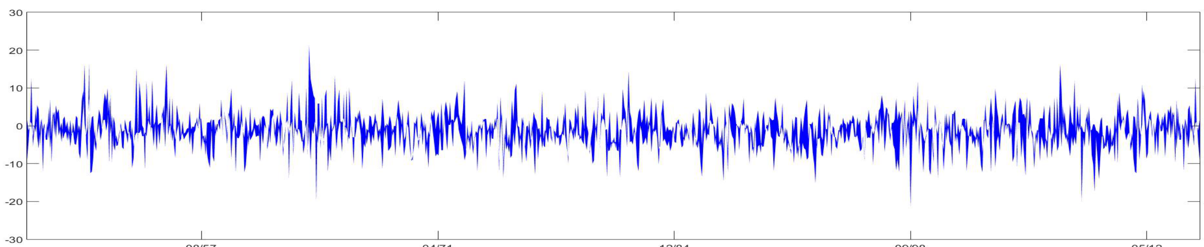
		Model Momentum (M.M.)				Residual Momentum (R.M.)			
Model	(K, L)	Mean	Vol.	S.R.	L.L.	Mean	Vol.	S.R.	L.L.
VAR-N	–	0.02	**5.0**	0.005	**-24.1**	**0.09**	5.8	0.015	-35.0
SV	–	**0.10**	**5.1**	0.019	-34.7	**0.11**	**5.6**	0.019	**-26.0**
VAR-SV	–	**0.12**	**4.5**	**0.028**	-20.2	**0.13**	5.8	**0.021**	-37.4
DFM-N	(1,1)	-0.04	**4.9**	-0.009	**-20.0**	**0.13**	**5.7**	**0.023**	-34.4
DFM-N	(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
DFM-SV	(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
DFM-SV2	(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
FAVAR-SV	(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
FAVAR-SV2	(4,2)	0.08	**5.7**	0.014	-32.3	0.03	**5.2**	0.005	-37.4

that the choice of the number of factors and lags in the factor models influences results strongly in this case. Further, mean returns of the some model and strategy combinations, such as FAVAR-SV and DFM-SV, are equal or higher than those of the S.M. strategy. In summary, there exist clear differences in the results between the two strategies: in general, more complex model structures such as DFM-SV and FAVAR-SV are good in combination with the M.M. strategy while using more simple model structures in combination with the R.M. strategy already leads to good results on mean returns. Apparently, using this latter strategy implies *a learning from past errors* and compensates for the lack of model complexity.

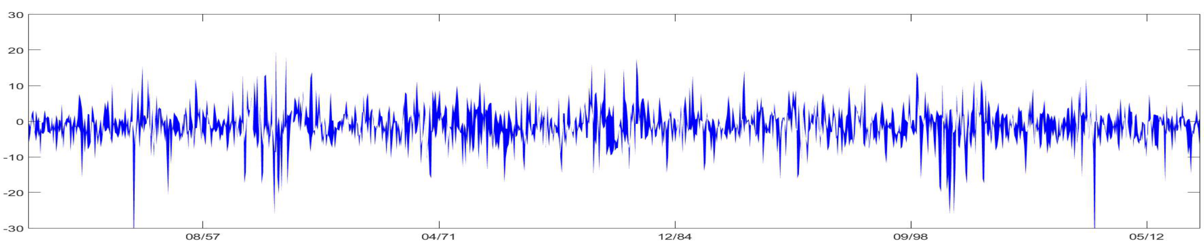
We next compare the volatility of realized returns in Table D.10. The differences between realized return volatilities between model and strategy combinations are less pronounced than the differences in mean realized returns. The obtained volatilities from each model and strategy combination are also close to that of the S.M. strategy. An interesting point is the comparison of model based and R.M. strategies. Given the same model class, M.M. generally leads to a lower volatility compared to R.M., but this difference is sensitive to the choice of the number of factors and lags in the factor models.

As additional indicators of risk, we report the Sharpe ratios and the largest loss from each model and strategy in Table D.10. Sharpe ratios, defined as scaled means and 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. 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. For all models, except SV, the largest loss is substantially lower for the M.M. strategy compared to R.M.. A complex model like the FAVAR-SV combined with M.M. leads to a very small extreme loss. Clearly, choice of momentum strategy matters substantially for risk of returns.

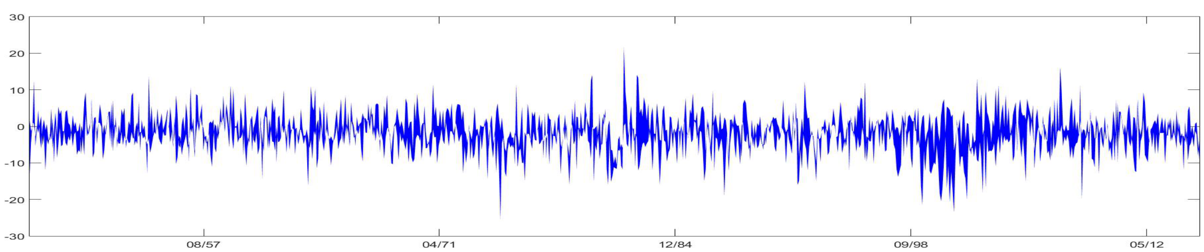
Figure D.8: *99% realized return intervals for three selected model and strategy combinations.*



(a) DFM (K = 1, L = 1) with M.M.



(b) DFM-SV (K = 3, L = 2) with R.M.



(c) DFM-SV (K = 4, L = 1) with M.M.

An important advantage of Bayesian inference is to provide complete distributions of the realized returns from the specific model and investment strategy combinations. Figure D.8 presents the 99% intervals of the realized returns for three selected model and portfolio strategies: DFM(1,1) with M.M., DFM-SV(4,1) with M.M. and DFM-SV(3,2) with R.M.. 99% intervals of returns are relatively tight for all three model and strategy combinations. In addition, even the worst-performing model and strategy combination, DFM(1,1) and M.M., has very high returns in some periods. Similarly, the best-performing strategies, DFM-SV(3,2) with R.M. and DFM-SV(4,1) with M.M. lead to extreme losses in some periods.[11] Apparently the time-variation in the performances of each model and strategy combination is important.

The realized return analysis of different model structures and investment strategies lead to two general conclusions. First, results for return and risk are sensitive to model and strategy choices, hence *diversity of results* should be taken into account. Second, several *return and risk features* are important: Using M.M. in combination with very simple models, like VAR-N and DFM-N that do not fit well, gives poor results. Complex models, like DFM-SV2 and FAVAR-SV2 tend to overfit and do not improve results compared to the models DFM-SV and FAVAR-SV. R.M. leads to reasonable returns for the simple models such as VAR-N, SV and DFM-N. However, it also does not improve the results for the complex DFM-SV2 and FAVAR-SV2 models compared to their simpler counterparts. M.M. strategy gives reasonable risk results for almost all models. There exists a sensitivity in the DFM class for the number of factors and lags. SV has poor risk. Using R.M. gives poor risk results for all models except for 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.

D.2 Return and risk features from combinations of model forecasts and investment strategies: additional results

This section presents additional results, particularly returns and risk from mixtures of models combined with a mixture of two investment strategies (M.M., R.M.). Realized returns and risk values of VAR-N, SV and DFM-N(4,2) combined with a mixture of M.M. and R.M. strategies presented in the main text. Similarly, realized returns and risk values of VAR-SV and DFM-SV(1-4, 1-2) models combined with two investment strategies (M.M. R.M.) are presented in the main text of the paper. In Table D.11, we present the detailed results for the third model and strategy combinations, namely FAVAR-SV models combined with a mixture of M.M. and R.M. strategies. For the last combination, we also present the mixture weights per number of factors in Figure D.9.

[11] These are general properties for all models we consider. We do not present all return intervals due to space considerations.

Table D.11: *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. S.M. 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 S.M.. 90% credible intervals are reported in parentheses.*

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

Figure D.9: *Weights (posterior mean) per combination 8 FAVAR-SV models and two strategies*