ΑΔΣ

Advances in Decision Sciences

Michael McAleer

Editor-in-Chief
University Chair Professor
Asia University, Taiwan

A Scoring Rule for Factor and Autoregressive Models Under Misspecification[*]

Roberto Casarin

University Ca' Foscari of Venice

Fausto Corradin

GRETA Associates

University Ca' Foscari of Venice

Francesco Ravazzolo

Free University of Bozen-Bolzano, Italy

CAMP, BI Norwegian Business School, Norway

Domenico Sartore[†]

University Ca' Foscari of Venice

Revised: June 2020

[†]Corresponding author: `sartore@unive.it`.

Abstract. Factor models (FM) are now widely used for forecasting with large set of time series. Another class of models, which can be easily estimated and used in a large dimensional setting, is multivariate autoregressive models (MAR), where independent autoregressive processes are assumed for the series in the panel. When applied to big data, the estimation, model selection and combination of both models can be time consuming. We assume both FM and MAR models are misspecified and provide a scoring rule which can be evaluated on an initial training sample to either select or combine the models in forecasting exercises on the whole sample. Some numerical illustrations are provided both on simulated data and on well known large economic datasets. The empirical results show that the frequency of the true positive signals is larger when FM and MAR forecasting performances differ substantially and it decreases as the horizon increases.

Keywords: Factor models, Large datasets, Multivariate autoregressive models, Forecasting, Scoring rules, VAR models.

JEL: C32, C52, C53.

1   Introduction

The recent fast growth in (real-time) big data allows researchers to model and predict variables of interest more accurately and suggests that there are large potential gains from using a big set of variables instead of a single univariate time series models in many inference applications. Consequently, new big databases have been developed and studied for economic problems, see, for example, Choi and Varian (2012); Varian (2014); Varian and Scott (2014); Einav and Levin (2014). Some leading examples in macroeconomics are Stock and Watson (2002, 2005); Bańbura et al. (2010); Koop and Korobilis (2013); Stock and Watson (2012). However, there are many issues still open when working with big datasets, including high-dimensional modeling and lack of inference efficiency. We refer to Granger (1998) for an early discussion of these issues, and to Litterman (1980); Sims and Zha (1998); Koop (2013) for a discussion from a Bayesian perspective.

Various papers advocated the use of regularization techniques to deal with the overparameterization and overfitting issues such as Lasso and Bayesian Lasso (Park and Casella, 2008; Hsu et al., 2008), adaptive Lasso, and elastic net and adaptive elastic net (Zou and Hastie, 2005; Zou and Zhang, 2009; Gefang, 2014). Other approaches consider stochastic search variable selection (George et al., 2008; Korobilis, 2013) graphical models (Ahelgebey et al., 2016a,b; Bianchi et al., 2019), random projections (Koop et al., 2017), Bayesian nonparametrics (Bassetti et al., 2014), Bayesian nonparametrics Lasso and spike-and-slab priors (Billio et al., 2019; Bassetti et al., 2020), hierarchical prior modelling (Billio et al., 2016). Other approaches rely on various forms of prior restrictions on the parameters (see, e.g. Koop, 2013; Koop and Korobilis, 2013; Korobilis, 2016). Finally, forecast combinations are also often applied to deal with large datasets, see, for example, (Stock and Watson, 2004; Raftery et al., 2010; Koop and Korobilis, 2012; Groen et al., 2013; Conflitti et al., 2015; Casarin et al., 2015).

In this paper we focus on two simple models that are widely used in forecasting: factor models (see, e.g. Stock and Watson, 2002, 2004, 2005; Bańbura et al., 2010; Stock and Watson, 2014, 2012), with reduced number of factors (FM), and multivariate autoregressive models (MAR), where no interaction is assumed between

the series in the panel, (e.g., see Penny and Harrison, 2006; Hytti et al., 2006). MAR models have been successfully used in various fields such as ecology (Ives et al., 2003) and neuroimaging (Harrison and Friston, 2003) and revealed their potential when applied to large set of time series. We refer to Gabriel Fagan (2006) for the use of time series models by the European Central Banks and to Reinsel (1983); Carriero et al. (2016); Cubadda and Guardabascio (2017) for extensions and recent applications to economics. In order to obtain some theoretical results and having in mind applications to very large set of time series, in this paper we will focus on a parsimonious specification of MAR. We assume a diagonal MAR model where lagged interactions among variables are excluded (Marcellino et al., 2006). The model is very simple and can be extended along various directions, for example by including, lagged interactions, exogenous covariates and interaction between covariates.

We derive a scoring rule for FM and MAR models. Various scoring rules have been proposed in the literature (Mitchell and Hall, 2005; Gneiting and Raftery, 2007; Gneiting and Ranjan, 2011, 2013; Lerch et al., 2016), but in order to preserve some analytical tractability of the comparison we consider mean square errors (MSEs) of the vectors of point forecasts generated with the two models. MSEs are the forecast error second moments and their trace is often used to evaluate the forecast accuracy. See Hendry and Martinez (2017) for a discussion and an alternative proposal in term of MSE determinants.

We expect that the use of a reduced number of factors in the FM model and of prior parameter restrictions in the MAR model, may lead to model misspecification. Thus, in this paper we follow some forecast analysis studies, see, e.g.,Schorfheide (2005), and score the two models under the assumption they are misspecified. Our new model-specific scoring rule for FM and MAR models indicates that the forecasting performances of the models depend crucially on the parameters setting of data generating process. More specifically, the goodness of the forecasting depends on the level of simultaneous or lagged dependence between the series. The proposed scoring rule is well suited for big data applications since it can be used on a initial set of observations to choose the forecasting model on the remaining set of data. It can be generalized to models that include autoregressive and factor models features, such as the multivariate index-augmented autoregression (MIAAR) models

of Cubadda and Guardabascio (2019). Also, the scoring rule can be applied to combine forecasts as well (e.g., see Bates and Granger, 1969; Diebold and Pauly, 1987, 1990; Geweke and Amisano, 2010). See also Kapetanios et al. (2015); Bassetti et al. (2018) for generalized combination schemes and Billio et al. (2013); Pettenuzzo and Ravazzolo (2016) for forecast combination based on time-varying weights.

We provide an illustration of the theoretical results and study the reliability of the proposed scoring rule through some simulation experiments and three applications to widely used datasets. The first database includes the quarterly Stock and Watson (2004) series; the second one the monthly McCracken and Ng (2015) series; the third one a set of cross-country series downloaded from Bloomberg which represent a three-country extension (EU, US and Japan) of the previous two datasets. We find that our scoring rule provides detects ex-ante the more accurate model in all these exercises. The reliability of the proposed scoring rule increases when accuracy differs substantially across models and it decreases as the horizon increases.

The paper is organized as follows. Section 2 introduces some notation and the models discussed in this paper. Section 3 derives the scoring rule. Section 4 applies the proposed scoring rule to simulation exercises. Section 5 exhibits some empirical results on well studied macroeconomic datasets. Section 6 concludes.

2   Forecasting models

We introduce some notation and define the factor (FM), vector autoregressive (VAR) and independent multivariate autoregressive (MAR) models used in this paper. Let $\{\mathbf{x}_t\}_{t \geq 0}$ be a n-dimensional real-value random process. In what follows, we assume the process is weak stationary and has zero mean and variance-covariance matrix $\mathbb{E}(\mathbf{x}_t\mathbf{x}_t') = \Gamma_X$ and auto-covariance function $\mathbb{E}(\mathbf{x}_t\mathbf{x}_{t+j}') = \Gamma_{X,j}$, $j \in \mathbb{Z}$ with $\Gamma_{X,j} = \Gamma_{X,-j}'$.

We denote with $\mathbf{a}_i$ and λ_i, $i = 1, \ldots, n$, the eigenvectors and eigenvalues, respectively of Γ_X such that

$$\Gamma_X \mathbf{a}_i = \lambda_i \mathbf{a}_i. \tag{1}$$

Let A be the orthogonal matrix with the normalized eigenvectors $\mathbf{a}_i$, $i = 1, \ldots, n$, in its columns and Λ the diagonal matrix with the eigenvalues λ_i, $i = 1, \ldots, n$ on the main diagonal, where we assume the eigenvalues are in a decreasing order, i.e. $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. It follows that the decomposition

$$\Gamma_X A = A\Lambda \tag{2}$$

holds true.

Under suitable conditions, the process $\{\mathbf{x}_t\}_{t \geq 0}$ can always be represented by using a set of factors $\mathbf{f}_{n,t} = (f_{1,t}, \ldots, f_{n,t})'$ and a factor loadings matrix A such that

$$\mathbf{x}_t = A\mathbf{f}_{n,t}, \quad t = 1, 2, \ldots, T. \tag{3}$$

In order to obtain a representation of $\{\mathbf{x}_t\}_{t \geq 0}$ on a lower dimensional space, we use the subset $\mathbf{f}_{k,t} = (f_{1,t}, \ldots, f_{k,t})'$ of the first k latent factors in $\mathbf{f}_{n,t}$, with $k < n$, and assume that the factors have an autoregressive dynamics. In our empirical applications the k factors are used to forecast $n^* < n$ variables of interest, whereas our theoretical results are presented for $n^* = n$, for the sake of simplicity and without loss of generality. Our FM model is defined by:

$$\mathbf{x}_t = A_k\mathbf{f}_{k,t} + \boldsymbol{\xi}_t \tag{4}$$

$$\mathbf{f}_{k,t} = \Phi_k\mathbf{f}_{k,t-1} + \boldsymbol{\eta}_{k,t}, \quad \boldsymbol{\eta}_{k,t} \sim \mathrm{WN}(\mathbf{0}, \Sigma_k) \tag{5}$$

$t = 1, \ldots, T$, where A_k is submatrix of A such that $A = (A_k | A_{n-k})$, $\{\boldsymbol{\xi}_t\}_{t \geq 1}$ is an idiosyncratic component with $\mathbb{E}(\boldsymbol{\xi}_t) = \mathbf{0}$, $\mathbb{C}ov(\boldsymbol{\xi}_t, \boldsymbol{\xi}_s) = \mathbf{0}$, $t \neq s$ and $\mathbb{V}(\boldsymbol{\xi}_t) = \Sigma_{\xi,t}$, and $\mathrm{WN}(\mathbf{0}, \Sigma_k)$ denotes a white noise process with mean $\mathbf{0}$ and variance-covariance matrix Σ_k. Furthermore, we assume the factors $\mathbf{f}_{k,t}$ admit the infinite MA representation

$$\mathbf{f}_{k,t} = \sum_{j=0}^{\infty} \Psi_{k,j}\boldsymbol{\eta}_{k,t-j}. \tag{6}$$

In the forecasting practice the factors $\mathbf{f}_{k,t}$ are first extracted and then predicted out of sample with a dynamic model (e.g. a VAR model). Predictions are then used

to generate forecasts for $\mathbf{x}_t$. The regression model $\mathbf{x}_t = (I_n \otimes \mathbf{f}'_{k,t})\boldsymbol{\beta}_k + \tilde{\boldsymbol{\xi}}_t$ is usually employed to recover relationship between dependent variables and the factors.

Remark 1. *We shall notice that this model does not allow for extracting further information from the data than the one encoded in the matrix A_k. The least square estimator $\hat{\boldsymbol{\beta}}_k$ of $\boldsymbol{\beta}_k$ is equal to $vec(A'_k)$ (see Appendix A). Thus the forecasting performance depends crucially on the predictability of the factors and the choice of the reduced number of factors k to use for forecasting the n^* variables of interest.*

The second forecasting model used in this paper is a MAR, which is defined as

$$\mathbf{x}_t = \Phi\mathbf{x}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathrm{WN}(\mathbf{0}, \Sigma_\eta) \tag{7}$$

$t = 1, 2, \ldots, T$, where $\Phi = \mathrm{diag}\{(\phi_{1,1}, \ldots, \phi_{1,n})'\}$ is a diagonal coefficient matrix and $\Sigma_\eta = \mathrm{diag}\{(\sigma_1^2, \ldots, \sigma_n^2)'\}$ is a diagonal variance-covariance matrix. We assume the process $\{\mathbf{x}_t\}_{t \geq 0}$ admits the infinite MA representation

$$\mathbf{x}_t = \sum_{j=0}^{\infty} \Psi_{m,j}\boldsymbol{\eta}_{t-j}. \tag{8}$$

We will compare the FM and MAR models under the assumption they are misspecified. We assume that data generating process (DGP) for $\{\mathbf{x}_t\}_{t \geq 0}$ comes from a VAR process of the first order with infinite MA representation

$$\mathbf{x}_t = \sum_{j=0}^{\infty} \Psi_{X,j}\boldsymbol{\varepsilon}_{t-j}, \quad \boldsymbol{\varepsilon}_t \sim \mathrm{WN}(\mathbf{0}, \Sigma_\varepsilon). \tag{9}$$

The difference between a MAR model and a VAR relates to the assumptions on lags order and on Σ_η. Appendix A describes the linkage between a factor model and a VAR(1) model.

3   Forecast accuracy

Let $\mathbf{x}_{j,t+s|t}$ denote the s-step-ahead forecast for $\mathbf{x}_{t+s}$ made at time t with a given model j, with $j \in \{k, m\}$, where k indicates the factor model and m the MAR

model. We define the mean square forecast error (MSE) as

$$MSE(s) = \mathbb{E}\left((\mathbf{x}_{t+s} - \mathbf{x}_{j,t+s|t})(\mathbf{x}_{t+s} - \mathbf{x}_{j,t+s|t})'\right), \quad s = 1, 2, \dots \tag{10}$$

The following theorems give the MSE for FM and MAR models and two useful decompositions, which will be used to evaluate the model's forecast accuracy. The MSE is derived under the assumption that the data generating process is a VAR process.

Theorem 1. *Assume* $\mathbf{x}_t$ *follows the model in Eq. 9. The MSE's trace for the FM can be decomposed as follows*

$$tr(MSE_k(s)) = tr\left(\mathbb{E}\left(\mathbf{e}_{k,t+s|t}\mathbf{e}'_{k,t+s|t}\right) + \sum_{j=0}^{s-1} \Psi_{X,j}\Sigma_\varepsilon\Psi'_{X,j}\right) \tag{11}$$

where $\mathbf{e}_{k,t+s|t} = \mathbf{x}_{t+s|t} - \mathbf{x}_{k,t+s|t}.$

Proof: see Appendix A.

Theorem 2. *The MSE's trace for the MAR model is*

$$tr(MSE_m(s)) = tr\left(\mathbb{E}\left(\mathbf{e}_{m,t+s|t}\mathbf{e}'_{m,t+s|t}\right)\right) + \sum_{j=1}^{n} \sigma_j^2\left(1 - \frac{\gamma_{j,1}^{2s}}{\sigma_j^{4s}}\right) \tag{12}$$

where $\mathbf{e}_{m,t+s|t} = \mathbf{x}_{t+s|t} - \mathbf{x}^*_{t+s|t}$ *and* $\mathbf{x}^*_{t+s|t}$ *is the forecast under the assumption of MAR dynamics and* $\gamma_{j,1}$ *is the* j-th *element of the main diagonal of the MAR first order autocorrelation matrix* $\Gamma_{1k} = diag\{(\gamma_{1,1}, \dots, \gamma_{n,1})'\}$ *and* σ_j^2 *is the* $j-$th *element of the main diagonal of the covariance matrix* $\Sigma_\eta = \mathbb{E}\left(\boldsymbol{\eta}_t\boldsymbol{\eta}'_t\right).$

Proof: see Appendix A.

The following properties of the factors $\mathbf{f}_{k,t}$ will be used to derive the main result of the paper. It can be shown that, under our assumption on the DGP, and the assumption on the factor loading matrix A, the sets of factors $\mathbf{f}_{n,t}$ and $\mathbf{f}_{k,t}$ satisfy the following conditions

1. $\mathbb{E}(\mathbf{f}_{n,t}\mathbf{f}'_{n,t}) = A'\Gamma_X A = \Lambda.$

8

2. $\mathbb{E}(\mathbf{f}_{k,t}\mathbf{f}'_{k,t}) = \Gamma_k$, where we defined $\Gamma_k = \text{diag}\{(\lambda_1, \ldots, \lambda_k)'\}$.

3. $\mathbb{E}(\mathbf{f}_{k,t}\mathbf{f}'_{k,t+j}) = \Gamma_{k,j}, \ \forall j \in \mathbb{Z}$.

4. $A'_k \Gamma_{X,1} A_k = \Gamma_{k,1}$ and $A'_k \Gamma_X A_k = \Gamma_k$.

We assume both FM and MAR models are misspecified and provide in the following theorem a scoring rule which is a function of the FM and MAR parameters. The rule can be computed on an initial training sample to either select or combine the models on the whole sample and in out-of-sample forecasting exercises. Also, the rule can be applied recursively in sequential forecasting exercises.

Theorem 3. *Let* $\mathbf{e}_{k,t+s|t} = \mathbf{x}_{t+s|t} - \mathbf{x}_{k,t+s|t}$, $\mathbf{e}_{m,t+s} = \mathbf{x}_{t+s} - \mathbf{x}^*_{t+s}$ *and* $\mathbf{x}^*_{t+s}$ *be the value of the process under the assumption of MAR dynamics. If the following inequality is satisfied:*

$$tr\left(\mathbb{E}\left(\mathbf{e}_{k,t+s|t}\mathbf{e}'_{k,t+s|t}\right)\right) - tr\left(\mathbb{E}\left(\mathbf{e}_{m,t+s|t}\mathbf{e}'_{m,t+s|t}\right)\right) \leq \tag{13}$$
$$tr\left((\Gamma_{n,1}\Gamma_n^{-1})^s\Gamma_n((\Gamma_{n,1}\Gamma_n^{-1})')^s)\right) - \sum_{j=1}^{n}\sigma_j^2\frac{\gamma_{j,1}^{2s}}{\sigma_j^{4s}}$$

then $tr(MSE_k(s)) \leq tr(MSE_m(s))$.

Proof: see Appendix A.

The inequality shows that in the presence of misspecification one model is not always superior to the other in terms of MSE and its forecasting performance crucially depends on the covariance and auto-covariance structures in the series. In the following sections, we show how to use the inequality to score the FM and MAR models or to combine their forecasts (see Billio et al. (2013)). Our empirical applications will show that the inequality can be successfully used in a context of large datasets.

4 Simulation results

We generate a dataset of 57 time series of 290 time observations each from a VAR model of the first order

$$\mathbf{x}_t = \Phi \mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \overset{iid}{\sim} \mathrm{WN}(0, \Sigma_\varepsilon) \tag{14}$$

We consider two sets of experiments. In the first set we study the effect of the correlation between idiosyncratic error terms on the reliability of the proposed scoring rule. We assume the variance-covariance matrix is parametrized in the variance and correlation parameters, $\varrho \in (-1, 1)$ and σ^2, respectively, as follows

$$\Sigma_\varepsilon = \sigma^2 (I_n + \varrho(\boldsymbol{\iota}\boldsymbol{\iota}' - I_n)) \tag{15}$$

where $\boldsymbol{\iota} = (1, \ldots, 1)'$ is the unit vector and I_n the identity matrix. As regards the coefficients matrix, we assume

$$\Phi = (X'X)^{-1}X'Y \tag{16}$$

where Y and X are two random matrices of dimension $T \times n$ with i.i.d. entries generated from a standard normal $\mathcal{N}(0, 1)$. Two experiment settings are considered: "weakly correlated noise" ($\varrho = 0.2$), and "strongly correlated noise" ($\varrho = 0.7$). In all settings $\sigma^2 = 3$. For each settings MAR and FM models have been estimated on 25 expanding window samples and forecasts generated for 12 steps ahead.

In the second set of experiments we study the effect of the lagged dependence between series on the reliability of our scoring rule. We assume the VAR coefficient matrix is parametrized in the variance and causality parameters, α and β, respectively, as follows

$$\Phi = (\alpha I_n + \beta(\boldsymbol{\iota}\boldsymbol{\iota}' - I_n)) \tag{17}$$

where α and β are such that the process is weak stationary. We consider the following settings: $\alpha = 0.01$ and $\alpha = 0.5$ with $\beta = (1 - \alpha - 0.01)/n$. The covariance matrix

Σ_ε is parametrized in ϱ and σ^2 as in the first set of experiments, with $\varrho = 0.9$ and $\sigma^2 = 3$.

We evaluate the trace of the mean square error, $\text{tr}(MSE_j(s,i))$, for all subsamples $i = 1, \ldots, 25$, horizons, $s = 1, \ldots, 12$, and models $j \in \{k, m\}$. For each pair (s, i), we compute the following scoring rule

$$C(s,i) = \text{tr}\left(\widehat{U}_k - \widehat{U}_m - (\hat{\Gamma}_{n,1}\hat{\Gamma}_n^{-1})^s \hat{\Gamma}_n((\hat{\Gamma}_{n,1}\hat{\Gamma}_n^{-1})')^s\right) + \sum_{j=1}^n \hat{\sigma}_j^2 \frac{\hat{\gamma}_{j,1}^{2s}}{\hat{\sigma}_j^{4s}}$$

obtained from the inequality in Th. 3, where we set

$$U_k = \mathbb{E}\left(\mathbf{e}_{k,t+s|t}\mathbf{e}'_{k,t+s|t}\right), \quad U_m = \mathbb{E}\left(\mathbf{e}_{m,t+s|t}\mathbf{e}'_{m,t+s|t}\right).$$

If $C < 0$ then FM is underperforming MAR. In our applications, evaluating $\widehat{U}_m$ and $\widehat{U}_k$ would require the estimation of a VAR on n variables, which might be not feasible for large n, thus forecasting error estimates are obtained with a VAR model on the n^* variables of interest.

We study the reliability of the proposed scoring rule by evaluating the frequency of the true positives and false negatives in our 25 samples. More specifically, we count the proportion of times the ordering of the models induced by the scoring rule C agrees with the one induced by their MSE, that is

$$f(s) = \frac{1}{25}\sum_{i=1}^{25} \mathbb{I}(\text{tr}(MSE_k(s,i)) < \text{tr}(MSE_m(s,i)))\mathbb{I}(C(s,i) < 0) \quad (18)$$

$$+\frac{1}{25}\sum_{i=1}^{25} \mathbb{I}(\text{tr}(MSE_k(s,i)) > \text{tr}(MSE_m(s,i)))\mathbb{I}(C(s,i) > 0).$$

We also compute the average performance

$$f = \frac{1}{12}\sum_{s=1}^{12} f(s). \quad (19)$$

In Fig. 1 we report the frequency, $f(s)$, (vertical axis) at different forecasting horizons, s, (horizontal axis) for the three datasets (different rows). Our results

11

show that the reliability of the scoring rule changes over the horizons and depends crucially on the dependence structure between the series. The left plot shows that larger simultaneous dependence, $\varrho = 0.7$, induces an increase in the reliability on the short-term horizons. In the right plot, one can see that larger lagged dependence levels, $\alpha = 0.5$, has impact on both long- and short-term horizons.

INSERT FIGURE 1 HERE

5   Empirical results

We consider some well known datasets used in macroeconomics. The first one is an extension of the Stock and Watson (2005) (SW dataset), which consists of 144 major macroeconomic time series sampled at quarterly frequency from 1959Q1 to 2011Q2. The dataset includes only revised series and not vintages of real-time data, when data are revised. The predictors include series in 14 categories: real output and income; employment and working hours; real retail, manufacturing and trade sales; consumption; housing starts and sales; real inventories; orders; stock prices; exchange rates; interest rates and spreads; money and credit quantity aggregates; price indexes; average hourly earnings; and miscellaneous. In order to deal with stationary series, we apply the series-specific transformation suggested in Stock and Watson (2005).[1]

The second dataset is the one described in McCracken and Ng (2015) (NM dataset). The dataset includes 120 time series sampled at monthly frequency from September 1992 to November 2016 and covers most of the time series used in the previous database. It exhibits some important appealing features. First, it is updated monthly using the FRED database. Second, it is publicly accessible in an easy manner, facilitating comparison of related research and replication of empirical work. McCracken and Ng (2015) show that factors extracted from this dataset share the similar predictive content as factors based on the Stock and Watson (2005) dataset.

---

[1]As in Stock and Watson (2005), we applied both outlier-adjusted and outlier-unadjusted versions of the series and findings are similar.

The third dataset used is a collection of 68 macroeconomics series from Bloomberg (BL dataset), sampled at monthly frequency from September 1993 to November 2017 for the US, EU and Japan. Precisely, it includes 31 variables for the US; 17 for EU; 18 for Japan; and the two exchange rates euro/yen and euro/US dollar. The list of variables is similar to the previous databases and contains different measures of core and headline prices; labor market variables; imports and exports; industrial production; consumption; sales; leading indicators; and several interest rates. See Tables 1-2 for the complete list of variables.

<div align="center">INSERT TABLES 1-2 HERE</div>

For each dataset we estimate the FM and MAR models on 25 sub-samples of increasing size. On each sub-sample we forecast n^* variables of interest (see Tab. 3) 12 steps ahead.

<div align="center">INSERT FIGURE 2 HERE</div>

The variables of interest of each dataset include both macroeconomic and financial indicators, that is:

SW: Industrial Production; Unemployment Rate; CPI-All Items; Fed Funds Market Rate.

NM: Real Personal Income, Industrial Production Index, Unemployment Rate, Effective Federal Funds, and CPI-All Items.

B: Personal Consumption expenditures (US), Harmonized Index of Consumer Price (EU), Consumer Price Index All Items (JP), Industrial Production(US,EU,JP), 2-Years Treasury (US), 2-Years Governative Bonds (EU, JP), and Unemployment Rate (US, EU, JP).

We report in Figure 3 the 25 sets of 12-step-ahead forecasts (red lines) for various US variables obtained by fitting FM (first line) and MAR (second line) models on the SW, NM and BL datasets. We report in Fig. 2 the sequences of generated forecasts from the two models for two variables of interest, industrial production and unemployment rate, for the twelve horizons and the three databases. Also, the forecasts for the EU and Japan variables are reported in Figure 4.

<div align="center">13</div>

INSERT FIGURES 3-4 HERE

In the FM we choose the number of factors (column k in Tab. 3) such that it is the same in all datasets and the percentage of variance explained is above the 75% (column V in Tab. 3). We evaluate the mean square error expressions given in Th. 1-2, $MSE_j(s,i)$, for all subsamples $i = 1, \ldots, 25$, horizons, $s = 1, \ldots, 12$, and models $j \in \{k, m\}$.

INSERT TABLE 3 HERE

The average MSE over samples and horizon is reported in Tab. 3, whereas the horizon-specific MSE averaged over samples is reported in the first and second column of Fig. 5.

INSERT FIGURE 5 HERE

The main evidence is that the FM performs on average less accurately than the MAR in all three databases. The difference varies across database: average MSE for the FM equal to 0.097 versus average MSE for MAR equal to 0.094 in the SW database; average MSE for the FM equal to 0.039 versus average MSE for MAR equal to 0.034 in the NM database; MSE for the FM equal to 0.021 versus average MSE for MAR equal to 0.017 in the B database. In the NM the loss is economically important and up to 8%; difference is smaller for the SW database where the two models perform similarly. Looking to Fig. 5, the MAR outperforms on average the FM mainly at short horizons in the SW and NM database. Maximum and minimum errors of the two models are more comparable over horizons. In the case of the Bloomberg database, MAR provides more accurate forecasts more uniformly across horizons. Therefore, our finding is similar to recent literature on the lack of superior forecast abilities of the FM models, when used for some variables or forecast horizons, see, for example, Medeiros et al. (2018).

In all three cases, predictions of MAR models are less volatile and flatter than those of the FM (e.g., see trending behaviour of the forecasts in Fig. 2). Despite being flatter, MAR predictions are more accurate. Selected factors explain a

large proportion of variance in the data, but this information does not translate automatically to accurate forecasts.

For each pair (s, i), we compute the scoring rule C and the frequency of the true positives and false negatives in our 25 samples at the 12 horizons, $f(s)$ and the average frequency across horizons, f (see column f in Tab. 3). In the third column of Fig. 5 we report the frequency (vertical axis) at different forecasting horizons (horizontal axis) for the three datasets (different rows). The scoring rule is re-computed at each vintage with only in-sample information up to the time of forecasting, therefore it is an ex-ante measure.

For two databases, MN and Bloomberg, the scoring rule at short and middle horizons has a success rate above 80%. Aastveit et al. (2018) document that a combination of factor models with different number of factors perform accurately and their methodology could be extended with the new scoring rule. The performance of the scoring rule decreases at longer horizons. For the SW, performance is weaker over all horizons, but we recall from Tab. 3 that FM and MAR perform quite similarly in this example, whereas MAR provides average forecasting gains in the other two exercises.

<center>INSERT FIGURE 6 HERE</center>

Another interesting empirical finding allows us to shed some light on the forecasting abilities of the factor models. The worse performance of FM with respect to MARs is strictly related to the low predictability of the factors in despite of their ability to explain a large proportion of variance of the panel of time series. As an example we report here the results for the Bloomberg dataset. Figure 6 shows the increase in the proportion of the explained variance when increasing the number of factors (dashed line, left axis). In the same figure, the total adjusted R^2 of the VAR regression on the factors (averaged over horizons and variables, solid circles line, left axis) decreases with the number of factors included, which is due to the loss of degrees of freedoms when including new variables in the VAR. The trade-off between loss of degrees of freedom and proportion of variance explained is reflected by the average MSEs (averaged over horizons and variables, solid lines, right axis). The MSEs for each variable are reported in Fig. 7.

<center>15</center>

At all forecasting horizons the average MSE (solid lines, right axis) do not decrease monotonically with the number of factors included. More precisely, after reaching a minimum value when 7 factors are used, the forecasting performances of the FM deteriorate with the number of factors. Interestingly, the increase in the MSEs is due to a decrease in the total adjusted R^2 (solid circles line) of the dynamic model on the factors and to a consequent loss of efficiency in forecasting the factors.

Our results provide also some guidelines for improving the predictive ability of FM models. In the forecasting regression, factors should be chosen specifically for each predicted variable, by applying some model selection procedures, or by estimating factor loading with regularizing techniques proposed in sparse factor and principal component analysis (e.g., see Carvalho et al., 2008; Zou et al., 2011; Qi et al., 2013; Lan et al., 2014; Ročková and George, 2016). Also horizon-specific models could be considered and the dynamic properties accounted when extracting the factors. We leave these issues for further research.

6   Conclusion

This paper establishes a set of conditions to be satisfied for a factor model to overperform a multivariate autoregressive model in terms of mean square forecasting error. The condition results in a scoring rule that can be used to select between the two models or to combine them. Furthermore, in the paper we show the performance of the scoring rule in simulation exercises where both the factor model and the multivariate autoregressive model are misspecified. Then, the analysis continues with three well-known macroeconomic datasets. The empirical results show that the frequency of the true positive signals is larger when factor model and the multivariate autoregressive model forecasting performances differ substantially. It also documents that factor models are not providing more accurate forecasts uniformly across variables and horizons.

References

Aastveit, K. A., Ravazzolo, F., and van Dijk, H. K. (2018). Combined density nowcasting in an uncertain economic environment. *Journal of Business Economics & Statistics*, 36(1):131–145.

Ahelgebey, D. F., Billio, M., and Casarin, R. (2016a). Bayesian Graphical Models for Structural Vector Autoregressive Processes. *Journal of Applied Econometrics*, 31(2):357–386.

Ahelgebey, D. F., Billio, M., and Casarin, R. (2016b). Sparse Graphical Vector Autoregression: A Bayesian Approach. *Annals of Economics and Statistics*, 123:333–361.

Bańbura, M., Giannone, D., and Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, 25:71–92.

Bassetti, F., Casarin, R., and Leisen, F. (2014). Beta-product dependent Pitman-Yor processes for Bayesian inference. *Journal of Econometrics*, 180(1):49–72.

Bassetti, F., Casarin, R., and Ravazzolo, F. (2018). Bayesian nonparametric calibration and combination of predictive distributions. *Journal of the American Statistical Association*, 1:1–30.

Bassetti, F., Casarin, R., and Rossini, L. (2020). Hierarchial hierarchical species sampling models. *Bayesian Analysis*.

Bates, J. M. and Granger, C. W. J. (1969). Combination of Forecasts. *Operational Research Quarterly*, 20:451–468.

Bianchi, D., Billio, M., Casarin, R., and Guidolin, M. (2019). Modeling systemic risk with markov switching graphical sur models. *Journal of Econometrics*, 210(1):58–74.

Billio, M., Casarin, R., Ravazzolo, F., and Van Dijk, H. (2016). Interactions between eurozone and US booms and busts: A Bayesian panel Markov-switching VAR model. *Journal of Applied Econometrics*, 31(7):1352–1370.

Billio, M., Casarin, R., Ravazzolo, F., and van Dijk, H. K. (2013). Time-varying combinations of predictive densities using nonlinear filtering. *Journal of Econometrics*, 177:213–232.

Billio, M., Casarin, R., and Rossini, L. (2019). Bayesian nonparametric sparse var models. *Journal of Econometrics*, 212(1):97–115.

Carriero, A., Kapetanios, G., and Marcellino, M. (2016). Structural analysis with multivariate autoregressive index models. *Journal of Econometrics*, 192(2):332–348.

Carvalho, C. M., Chang, J., Lucas, J. E., Nevins, J. R., Wang, Q., and West, M. (2008). High-dimensional sparse factor modeling: Applications in gene expression genomics. *Journal of the American Statistical Association*, 103(484):1438–1456.

Casarin, R., Grassi, S., Ravazzolo, F., and Van Dijk, H. (2015). Dynamic predictive density combinations for large data sets in economics and finance. Technical report, Tinbergen Institute Discussion Paper 15-084/III.

Choi, H. and Varian, H. (2012). Predicting the present with Google trends. *Economic Record*, 88:2–9.

Conflitti, C., De Mol, C., and Giannone, D. (2015). Optimal combination of survey forecasts. *International Journal of Forecasting*, 31(4):1096–1103.

Cubadda, G. and Guardabascio, B. (2017). Representation, estimation and forecasting of the multivariate index-augmented autoregressive mode. Technical report, CEIS Research Paper 397, Tor Vergata University.

Cubadda, G. and Guardabascio, B. (2019). Optimal combination of survey forecasts. *Representation, estimation and forecasting of the multivariate index-augmented autoregressive model*, 35(7):67–79.

Diebold, F. and Pauly, P. (1990). The use of prior information in forecast combination. *International Journal of Forecasting*, 6:503–508.

Diebold, F. X. and Pauly, P. (1987). Structural change and the combination of forecasts. *Journal of Forecasting*, 6:21–40.

Einav, L. and Levin, J. (2014). Economics in the age of big data. *Science*, 346(6210):715–718.

Gabriel Fagan, J. M. (2006). *Econometric Models of the Euro-area Central Banks*. Elgar.

Gefang, D. (2014). Bayesian doubly adaptive elastic-net Lasso for VAR shrinkage. *International Journal of Forecasting*, 30(30):1–11.

George, E. I., Sun, D., and Ni, S. (2008). Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics*, 142(1):553–580.

Geweke, J. and Amisano, G. (2010). Optimal prediction pools. *Journal of Econometrics*, 164(2):130–141.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102:359–378.

Gneiting, T. and Ranjan, R. (2011). Comparing density forecasts using threshold and quantile weighted scoring rules. *Journal of Business and Economic Statistics*, 29:411–422.

Gneiting, T. and Ranjan, R. (2013). Combining predicitve distributions. *Electronic Journal of Statistics*, 7:1747–1782.

Granger, C. W. J. (1998). Extracting information from mega-panels and high-frequency data. *Statistica Neerlandica*, 52:258–272.

Groen, J. J. J., Paap, R., and Ravazzolo, F. (2013). Real-time inflation forecasting in a changing world. *Journal of Business & Economic Stastistics*, 31:29–44.

Harrison, L. anad Penny, W. D. and Friston, K. (2003). Multivariate autoregressive modeling of fmri time series. *NeuroImage*, 19:1477–1491.

Hendry, D. F. and Martinez, A. B. (2017). Evaluating multi-step system forecasts with relatively few forecast-error observations. *International Journal of Forecasting*, 33(2):359 – 372.

Hsu, N.-J., Hung, H.-L., and Chang, Y.-M. (2008). Subset selection for vector autoregressive processes using lasso. *Computational Statistics and Data Analysis*, 52:3645–3657.

Hytti, H., Takalo, R., and Ihalainen, H. (2006). Tutorial on multivariate autoregressive modelling. *Journal of Clinical Monitoring and Computing*, 20(2):101–108.

Ives, A. R., Dennis, B., Cottingham, K. L., and Carpenter, S. R. (2003). Estimating community stability and ecological interactions from time-series data. *Ecological Monographs*, 301-330:73.

Kapetanios, G., Mitchell, J., Price, S., and Fawcett, N. (2015). Generalised density forecast combinations. *Journal of Econometrics*, 188:150–165.

Koop, G. (2013). Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics*, 28(2):177–203.

Koop, G. and Korobilis, D. (2012). Forecasting inflation using dynamic model averaging. *International Economic Review*, 53:867–886.

Koop, G. and Korobilis, D. (2013). Large time-varying parameter VARs. *Journal of Econometrics*, 177:185–198.

Koop, G., Korobilis, D., and Pettenuzzo, D. (2017). Bayesian compressed VARs. *Journal of Econometrics*, 1:1–30.

Korobilis, D. (2013). VAR forecasting using Bayesian variable selection. *Journal of Applied Econometrics*, 28(2):204–230.

Korobilis, D. (2016). Prior selection for panel vector autoregressions. *Computational Statistics & Data Analysis*, 101:110–120.

Lan, A. S., Waters, A. E., Studer, C., and Baraniuk, R. G. (2014). Sparse factor analysis for learning and content analytics. *Journal of Machine Learning Research*, 15:1959–2008.

Lerch, S., Thorarinsdottir, T., Ravazzolo, R., and Gneiting, T. (2016). Forecaster's dilemma: Extreme events and forecast evaluation. *Statistical Science*, 32(1):106–127.

Litterman, R. (1980). Techniques for forecasting with vector autoregressions. *University of Minnesota, Ph.D. Dissertation*.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer, Berlin.

Marcellino, M., Stock, J. H., and Watson, M. W. (2006). A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series. *Journal of Econometrics*, 135(1):499 – 526.

McCracken, M. W. and Ng, S. (2015). FRED-MD: A monthly database for macroeconomic research. *NBER Working paper 2015-012*.

Medeiros, M., Veiga, A., Vasconcelos, G., and Zilberman, E. (2018). Forecasting inflation in a data-rich enviroment: the benefits of machine learning methods. Technical report, SSRN.

Mitchell, J. and Hall, S. G. (2005). Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESER "fan" charts of inflation. *Oxford Bulletin of Economics and Statistics*, 67:995–1033.

Park, T. and Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, 103(482):681–686.

Penny, W. and Harrison, L. (2006). Chapter 40: Multivariate autoregressive models.

Pettenuzzo, D. and Ravazzolo, F. (2016). Optimal portfolio choice under decision-based model combinations. *Journal of Applied Econometrics*, 31(7):1312–1332.

Qi, X., Luo, R., and Zhao, H. (2013). Sparse principal component analysis by choice of norm. *Journal of Multivariate Analysis*, 114:127 – 160.

Raftery, A. E., Kárńy, M., and Ettler, P. (2010). Online prediction under model uncertainty via Dynamic Model Averaging: Application to a cold rolling mill. *Technometrics*, 52:52–66.

Reinsel, G. (1983). Some results on multivariate autoregressive index models. *Biometrika*, 70(1):145–156.

Ročkovà, V. and George, E. I. (2016). Fast bayesian factor analysis via automatic rotations to sparsity. *Journal of the American Statistical Association*, 111(516):1608–1622.

Schorfheide, F. (2005). VAR forecasting under misspecification. *Journal of Econometrics*, 128(1):99 – 136.

Sims, C. A. and Zha, T. (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, 39(4):949–968.

Stock, J. H. and Watson, W. M. (2002). Forecasting using principal components from a large number of predictors. *Journal of American Statistical Association*, 97:1167–1179.

Stock, J. H. and Watson, W. M. (2004). Combination forecasts of output growth in a seven - country data set. *Journal of Forecasting*, 23:405–430.

Stock, J. H. and Watson, W. M. (2005). Implications of dynamic factor models for VAR analysis. Technical report, NBER Working Paper No. 11467.

Stock, J. H. and Watson, W. M. (2012). Disentangling the channels of the 2007-09 recession. *Brookings Papers on Economic Activity*, pages 81–156, Spring.

Stock, J. H. and Watson, W. M. (2014). Estimating turning points using large data sets. *Journal of Econometris*, 178:368–381.

Varian, H. (2014). Machine learning: New tricks for econometrics. *Journal of Economics Perspectives*, 28:3–28.

Varian, H. and Scott, S. (2014). Predicting the present with Bayesian structural time series. *International Journal of Mathematical Modelling and Numerical Optimisation*, 5:4–23.

Zou, H., Choi, J., and Oehlert, G. (2011). A penalized maximum likelihood approach to sparse factor analysis. *Statistics and its Interface*, 3:429–436.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67:301–320.

Zou, H. and Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. *Annals of Statistics*, 37:1733–1751.

A   Proof of the results in the paper

A.1   Proof of the result in Remark 1

Let X be the $T \times n$ sample matrix and A_k the matrix with the first k orthonormal vectors of X, and F_k the associated $T \times k$ factor matrix such that

$$F_k = X A_k \tag{A.1}$$

The least square estimator $\hat{B}_k$ of B_k in the regression model

$$X = F_k B_k + \boldsymbol{\xi} \tag{A.2}$$

is $\hat{B}_k = (F_k' F_k)^{-1} F_k' X$ which satisfies

$$F_k' F_k \hat{B}_k = F_k' F_k (F_k' F_k)^{-1} F_k' X = F_k' X \tag{A.3}$$

from Eq. A.1 and since $(F_k' F_k)^{-1} = \Lambda_k^{-1}$ one obtains

$$\hat{B}_k = \Lambda_k^{-1} A_k' X' X \tag{A.4}$$

which implies $\hat{B}_k = A_k'$ where A_k satisfies Eq. A.1

A.2   Linkage factor model and VAR(1) model

Let $\mathbf{x}_t$ be a n-dimensional real-value vector, with zero mean and stationary variance-covariance matrix. Then:

$$\mathbb{E}(\mathbf{x}_t \mathbf{x}_t') \mathbf{a}_i = \Gamma_X \mathbf{a}_i = \lambda_i \mathbf{a}_i \tag{A.5}$$

where $\mathbf{a}_i$ and λ_i, $i = 1, \ldots, n$ are the eigenvectors and eigenvalues of Γ_X and A is the $(n \times n)$ orthogonal matrix with the normalized eigenvectors $\mathbf{a}_i$, $i = 1, \ldots, n$, in its columns. Then, it follows that

$$\Gamma_X A = A \Lambda \tag{A.6}$$

with Λ the diagonal matrix with the eigenvalues λ_i, $i = 1, \ldots, n$. It follows that $A' \Gamma_X A = \Lambda$; $\Gamma_X = A \Lambda A'$.

The factors $\mathbf{f}_{n,t}$ are computed as

$$\mathbf{f}_{n,t} = A' \mathbf{x}_t \tag{A.7}$$

and the kth factor as

$$\mathbf{f}_{k,t} = A_k^{'}\mathbf{x}_t \tag{A.8}$$

where A_k is the matrix of the first k column of A and $A^{'}A = I$, $A_k^{'}A_k = I_k$. Therefore, we can compute

$$\tilde{\mathbf{x}}_{k,t} = A_k\mathbf{f}_{k,t} \tag{A.9}$$

which can proxy the true $\mathbf{x}_{k,t}$. Let now assume that $\mathbf{x}_t$ is generated by a VAR(1) model:

$$\mathbf{x}_t = \Phi_{\mathbf{x}}\mathbf{x}_{t-1} + \epsilon_{\mathbf{x,t}} \quad \epsilon_{\mathbf{x,t}} \sim i.i.d.N(0, \Sigma_\epsilon) \tag{A.10}$$

By pre-multiplying for $A_k^{'}$, we have:

$$A_k^{'}\mathbf{x}_t = A_k^{'}\Phi_{\mathbf{x}}\mathbf{x}_{t-1} + A_k^{'}\epsilon_{\mathbf{x,t}} \tag{A.11}$$

Fixing $\epsilon_{\mathbf{f,t}} = A_k^{'}\epsilon_{\mathbf{x,t}}$ and from (A.7)

$$\mathbf{f}_{k,t} = A_k^{'}\Phi_{\mathbf{x}}\mathbf{x}_{t-1} + \epsilon_{\mathbf{f,t}} \tag{A.12}$$

Then, we introduce a_k such as

$$A_k^{'}\Phi_{\mathbf{x}} = a_k^{'}A_k^{'} \tag{A.13}$$

Let multiply for A_k

$$A_k^{'}\Phi_{\mathbf{x}}A_k = a_k^{'}A_k^{'}A_k = a_k^{'} \tag{A.14}$$

since $A_k^{'}A_k = I_k$. Then,

$$\mathbf{f}_{k,t} = a_k^{'}\mathbf{f}_{k,t-1} + \epsilon_{\mathbf{f,t}} = A_k^{'}\Phi_{\mathbf{x}}A_k\mathbf{f}_{k,t-1} + \epsilon_{\mathbf{k,t}} \tag{A.15}$$

By pre-multiplying for A_k and recalling $\tilde{\mathbf{x}}_{k,t} = A_k\mathbf{f}_{k,t}$, we find the VAR(1) representation:

$$\tilde{\mathbf{x}}_{k,t} = A_kA_k^{'}\Phi_{\mathbf{x}}\tilde{\mathbf{x}}_{k,t-1} + A_kA_k^{'}\epsilon_{\mathbf{x,t}} \tag{A.16}$$

A.3   Proof of the result in Theorem 1

Let $\mathbf{x}_{t+s|t}$ denote the best linear forecast under the DGP, that is

$$\mathbf{x}_{t+s|t} = \sum_{j=0}^{\infty} \Psi_{s+j}\boldsymbol{\varepsilon}_{t-j}, \tag{A.17}$$

24

then by applying the infinite MA representation of $\mathbf{f}_{k,t+s}$ we obtain

$$\mathbf{x}_{k,t+s} = A_k \mathbf{f}_{k,t+s} = \sum_{j=0}^{\infty} A_k \Psi_{k,j} A_k' A_k \boldsymbol{\eta}_{t+s-j} \tag{A.18}$$

From the decomposition

$$\mathbf{x}_{t+s} - \mathbf{x}_{k,t+s|t} = \mathbf{e}_{t+s|t} + \mathbf{e}_{k,t+s|t} \tag{A.19}$$

where $\mathbf{e}_{t+s|t} = \mathbf{x}_{t+s} - \mathbf{x}_{t+s|t}$, $\mathbf{e}_{k,t+s|t} = \mathbf{x}_{t+s|t} - \mathbf{x}_{k,t+s|t}$, it follows that

$$
\begin{aligned}
MSE_k(s) &= \mathbb{E}\left((\mathbf{e}_{t+s|t} + \mathbf{e}_{k,t+s|t})(\mathbf{e}_{t+s|t} + \mathbf{e}_{k,t+s|t})'\right) \\
&= \mathbb{E}\left(\mathbf{e}_{t+s|t}\mathbf{e}_{t+s|t}'\right) + \mathbb{E}\left(\mathbf{e}_{k,t+s|t}\mathbf{e}_{k,t+s|t}'\right) \\
&= \mathbb{E}\left(\sum_{j=0}^{s-1} \Psi_{X,j}\boldsymbol{\varepsilon}_{t+s-j}\boldsymbol{\varepsilon}_{t+s-j}'\Psi_{X,j}'\right) + \mathbb{E}\left(\mathbf{e}_{k,t+s|t}\mathbf{e}_{k,t+s|t}'\right) \\
&= \sum_{j=0}^{s-1} \Psi_{X,j}\Sigma_\varepsilon \Psi_{X,j}' + \mathbb{E}\left(\mathbf{e}_{k,t+s|t}\mathbf{e}_{k,t+s|t}'\right)
\end{aligned}
$$

A.4   Proof of the result in Theorem 2

Decompose the MSE as follows

$$
\begin{aligned}
MSE_m(s) &= \mathbb{E}\left((\mathbf{x}_{t+s} - \mathbf{x}_{m,t+s|t})(\mathbf{x}_{t+s} - \mathbf{x}_{m,t+s|t})'\right) \tag{A.20} \\
&= \mathbb{E}\left((\mathbf{x}_{t+s} - \mathbf{x}_{t+s}^*)(\mathbf{x}_{t+s} - \mathbf{x}_{t+s}^*)'\right) + \mathbb{E}\left((\mathbf{x}_{t+s}^* - \mathbf{x}_{m,t+s|t})(\mathbf{x}_{t+s}^* - \mathbf{x}_{m,t+s|t})'\right)
\end{aligned}
$$

where $\mathbf{x}_{m,t+s|t} = \Phi \mathbf{x}_{m,t+s-1|t} = \Phi^s \mathbf{x}_t$. Assume $\mathbb{E}(\boldsymbol{\eta}_{t+1}\mathbf{x}_{m,t}') = \mathbb{E}((\mathbf{x}_{m,t+1} - \Phi \mathbf{x}_{m,t})\mathbf{x}_{m,t}') = 0$, as in Lütkepohl (2005), p. 34, then

$$\Phi = \mathbb{E}(\mathbf{x}_{m,t}\mathbf{x}_{m,t-1}')\mathbb{E}(\mathbf{x}_{m,t-1}\mathbf{x}_{m,t-1}')^{-1} = \Gamma_1 \Sigma_\eta^{-1} \tag{A.21}$$

Recall that $\mathbb{E}(\boldsymbol{\eta}_{k,t+1}\boldsymbol{\eta}_{k,t+1}') = \Sigma_k$, then

$$\Sigma_k = \Gamma_k - \Gamma_{1k}\Gamma_k^{-1}\Gamma_{-1k} = \Gamma_k - \Gamma_{1k}\Gamma_k^{-1}\Gamma_{1k}' \tag{A.22}$$

where $\Gamma_{1k} = \mathbb{E}\left(\mathbf{f}_{k,t}\mathbf{f}_{k,t-1}'\right) = \mathrm{diag}\{(\gamma_{1,1}, \ldots, \gamma_{k,1})'\}$ is the first order autocovariance matrix and $\Gamma_k = \mathbb{E}\left(\mathbf{f}_{k,t}\mathbf{f}_{k,t}'\right)$ the covariance matrix. By using Eq. A.21 and A.22 with $k = n$, we conclude that, the second term of the decomposition in Eq. A.20

can be written as

$$
\mathrm{tr}\left(\mathbb{E}\left((\mathbf{x}^*_{t+s} - \mathbf{x}_{m,t+s|t})(\mathbf{x}^*_{t+s} - \mathbf{x}_{m,t+s|t})'\right)\right) = \sum_{i=1}^{n}\sum_{h=0}^{s-1}\left(\sigma_i^2 - \frac{\gamma_{i,1}^2}{\sigma_i^2}\phi_i^h\right)\text{(A.23)}
$$

$$
= \sum_{j=1}^{n}\left(\sigma_j^2 - \frac{\gamma_{j,1}^2}{\sigma_j^2}\right)\frac{1-\phi_j^{2s}}{1-\phi_j^2} = \sum_{j=1}^{n}\sigma_j^2 - \sum_{j=1}^{n}\sigma_j^2\frac{\gamma_{j,1}^{2s}}{\sigma_j^{4s}}.
$$

A.5 Proof of the result in Theorem 3

From the properties of the factors given in Section 3, it follows that $\Phi_k = \mathbb{E}(\mathbf{f}_{k,t}\mathbf{f}'_{k,t-1})\mathbb{E}(\mathbf{f}_{k,t-1}\mathbf{f}'_{k,t-1})^{-1} = \Gamma_{k,1}\Gamma_k^{-1}$ and $\Sigma_k = \mathbb{E}((\mathbf{f}_{k,t}-\Phi_k\mathbf{f}_{k,t-1})(\mathbf{f}_{k,t}-\Phi_k\mathbf{f}_{k,t-1})') = \Gamma_k - \Gamma_{k,1}\Gamma_k^{-1}\Gamma_{k,1}$ and the factor mean forecasting error can be written as

$$
\mathbb{E}((\mathbf{f}_{t+s} - \mathbf{f}_{t+s|t})(\mathbf{f}_{t+s} - \mathbf{f}_{t+s|t})') = \sum_{j=0}^{s-1}\Psi_{k,j}\Sigma_k\Psi'_{k,j}
$$

$$
= \sum_{j=0}^{s}(\Gamma_{k,1}\Gamma_k^{-1})^j(\Gamma_k - \Gamma_{k,1}\Gamma_k^{-1}\Gamma'_{k,1})((\Gamma_{k,1}\Gamma_k^{-1})^j)'
$$

$$
= \Gamma_k + \sum_{j=1}^{s}(\Gamma_{k,1}\Gamma_k^{-1})^{j-1}\Gamma_k^{-1}((\Gamma_{k,1}\Gamma_k^{-1})^{j-1})' - \sum_{j=0}^{s}(\Gamma_{k,1}\Gamma_k^{-1})^j\Gamma_{k,1}\Gamma_k^{-1}\Gamma'_{k,1}((\Gamma_{k,1}\Gamma_k^{-1})^j)'
$$

$$
= \Gamma_k - (\Gamma_{k,1}\Gamma_k^{-1})^{s-1}\Gamma_{k,1}\Gamma_k^{-1}\Gamma'_{k,1}((\Gamma_{k,1}\Gamma_k^{-1})^{s-1})'
$$

$$
= \Gamma_k - ((\Gamma_{k,1}\Gamma_k^{-1})^{s-1}\Gamma_{k,1})\Gamma_k^{-1}((\Gamma_{k,1}\Gamma_k^{-1})^{s-1}\Gamma_{k,1})'
$$

$$
= \Gamma_k - (\Gamma_{k,1}\Gamma_k^{-1})^s\Gamma_k((\Gamma_{k,1}\Gamma_k^{-1})^s)'.
$$

It follows that

$$
\mathbb{E}((A_k\mathbf{f}_{t+s} - A_k\mathbf{f}_{t+s|t})(A_k\mathbf{f}_{t+s} - A_k\mathbf{f}_{t+s|t})') = \sum_{j=0}^{s-1}A_k\Psi_{k,j}\Sigma_k\Psi'_{k,j}A'_k
$$

$$
= A_k\left(\Gamma_k - (\Gamma_{k,1}\Gamma_k^{-1})^s\Gamma_k(\Gamma_{k,1}\Gamma_k^{-1})'^s\right)A'_k.
$$

By applying the trace operator one obtains

$$
\mathrm{tr}\left(A_k\left(\Gamma_k - (\Gamma_{k,1}\Gamma_k^{-1})^s\Gamma_k(\Gamma_{k,1}\Gamma_k^{-1})'^s\right)A'_k\right) = \mathrm{tr}\left(\Gamma_k - (\Gamma_{k,1}\Gamma_k^{-1})^s\Gamma_k(\Gamma_{k,1}\Gamma_k^{-1})'^s\right)
$$

$$
= \sum_{i=1}^{k}\lambda_i - \mathrm{tr}\left((\Gamma_{k,1}\Gamma_k^{-1})^s\Gamma_k(\Gamma_{k,1}\Gamma_k^{-1})'^s\right).
$$

since $A_k' A_k = I_k$. Considering that for $k = n$

$$\sum_{i=1}^{n} \lambda_i - \mathrm{tr}\left((\Gamma_{n,1}\Gamma_n^{-1})^s \Gamma_n (\Gamma_{n,1}\Gamma_n^{-1})^{*s}\right) = \sum_{j=0}^{s-1} \Psi_{X,j} \Sigma_\varepsilon \Psi_{X,j}'$$

and using the decomposition in Th. 1

$$\mathrm{tr}(\mathrm{MSE}_k(s)) = \mathrm{tr}\left(\mathbb{E}\left(\mathbf{e}_{k,t+s|t}\mathbf{e}_{k,t+s|t}'\right)\right) + \sum_{j=1}^{n} \lambda_j - \mathrm{tr}\left((\Gamma_{n,1}\Gamma_n^{-1})^s \Gamma_n ((\Gamma_{n,1}\Gamma_n^{-1})^s)'\right).$$

Finally, since $\sum_{j=1}^{n} \sigma_j^2 = \sum_{j=1}^{n} \lambda_j$ and applying the decomposition in Th. 2 we obtain the inequality:

$$\mathrm{tr}\left(\mathbb{E}\left(\mathbf{e}_{k,t+s|t}\mathbf{e}_{k,t+s|t}'\right)\right) - \mathrm{tr}\left(\mathbb{E}\left(\mathbf{e}_{m,t+s|t}\mathbf{e}_{m,t+s|t}'\right)\right)$$

$$\leq \mathrm{tr}\left((\Gamma_{n,1}\Gamma_n^{-1})^s \Gamma_n ((\Gamma_{n,1}\Gamma_n^{-1})')^s)\right) - \sum_{j=1}^{n} \sigma_j^2 \frac{\gamma_{j,1}^{2s}}{\sigma_j^{4s}}.$$

if and only if $\mathrm{tr}(\mathrm{MSE}_k(s)) \leq \mathrm{tr}(\mathrm{MSE}_m(s))$.

Table 1: Bloomberg data

Japan		Exch. rates	
Var.	Meas.	Var.	Meas.
Exports	m/m	EUJP	Change Ratio
Imports	m/m	EUUS	Change Ratio
Unemployment	%		
CPI	a/a		
CPI core	a/a		
PPI	a/a		
Industrial production	a/a		
Industrial machinery	a/a		
PMI manufactoring	level		
Private consumption	m/m		
Real income	m/m		
Retail sales	m/m		
Consumer confidence	level		
2 years yield	%		
3 years yield	%		
5 years yield	%		
7 years yield	%		
10 years yield	%		

Table 2: Bloomberg data

| US		EU	
Var.	Meas.	Var.	Meas.
Exports	m/m	Exports	m/m
Imports	m/m	Imports	m/m
Employment	%	Unemployment	%
Employment (agriculture)	thous	HCPI	a/a
Employment (private sector)	thous	CPI core	a/a
Average wages	m/m	PPI	a/a
PCE	a/a	Industrial production	a/a
PCE core	a/a	Construction	m/m
PPI	a/a	PMI manufactoring	level
PPI core	a/a	ESI	level
Industrial production	level	Leading indicator	level
Industrial orders	m/m	Retail sales	m/m
Industrial orders (durables)	m/m	2 years yield	%
Industrial orders (durables excl. transports)	m/m	3 years yield	%
Stocks	m/m	5 years yield	%
Capacity	%	7 years yield	%
ISM (manufacturing)	level	10 years yield	%
New buildings	m/m		
Construction spending	m/m		
Existing home sales	m/m		
New home sales	m/m		
Government spending	m/m		
Real income	m/m		
Conference Board index	level		
Michigan index	level		
2 years yield	%		
3 years yield	%		
5 years yield	%		
7 years yield	%		
10 years yield	%		
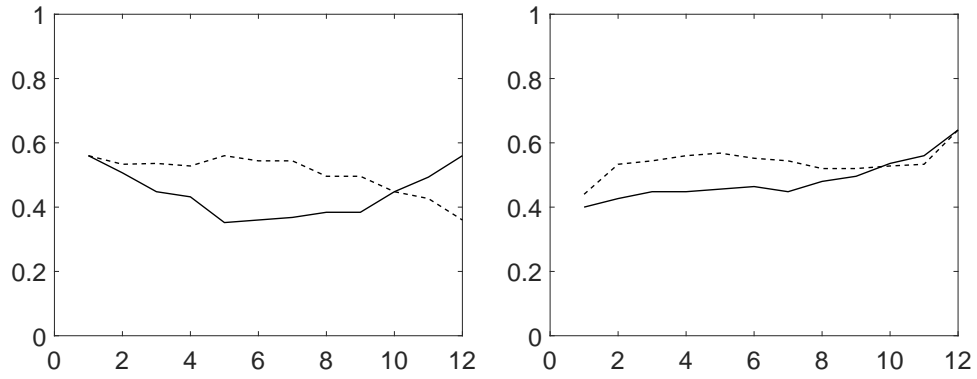
Figure 1: Scoring rule performance, $f(s)$, measured as the frequency (vertical axis) of a true positives and false negatives signals, over different horizons $s = 1, \ldots, 12$ (horizontal axis). Left: first experiments set with $\varrho = 0.2$ (solid), $\varrho = 0.7$ (dashed) and $\sigma^2 = 3$. Right: second experiments set with $\alpha = 0.01$ (solid), $\alpha = 0.5$ (dashed), and $\varrho = 0.7$ and $\sigma^2 = 3$.

Data	T	n	n^*	k	V	f	FM MSE	MAR MSE
SW	170	144	4	17	0.807	0.337	0.097	0.094
NM	290	120	5	17	0.745	0.827	0.039	0.034
B	290	68	12	17	0.807	0.870	0.021	0.017

Table 3: Comparison between FM and MAR models on three datasets: SW, NM and BL. T: number of observations; n: number of series; n^*: number of forecasted variables; k: number of factors; V: proportion of variance explained by factors; f: frequency of correctly predicted model orderings (true positives and false negatives); MSE: average mean square error.

(a) Industrial Production
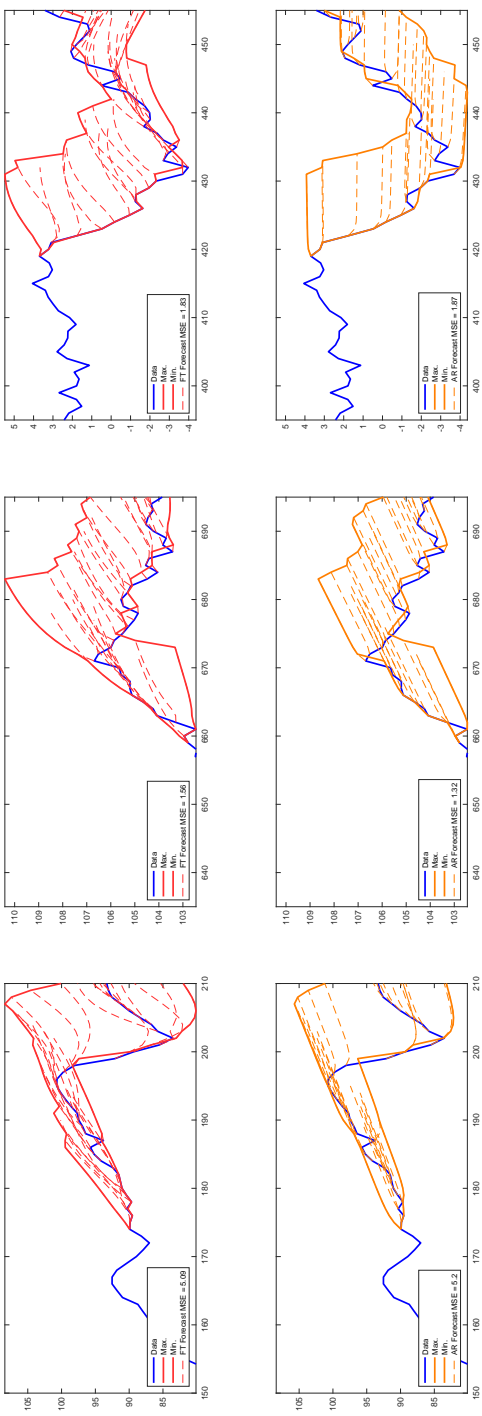
(b) Unemployment rate
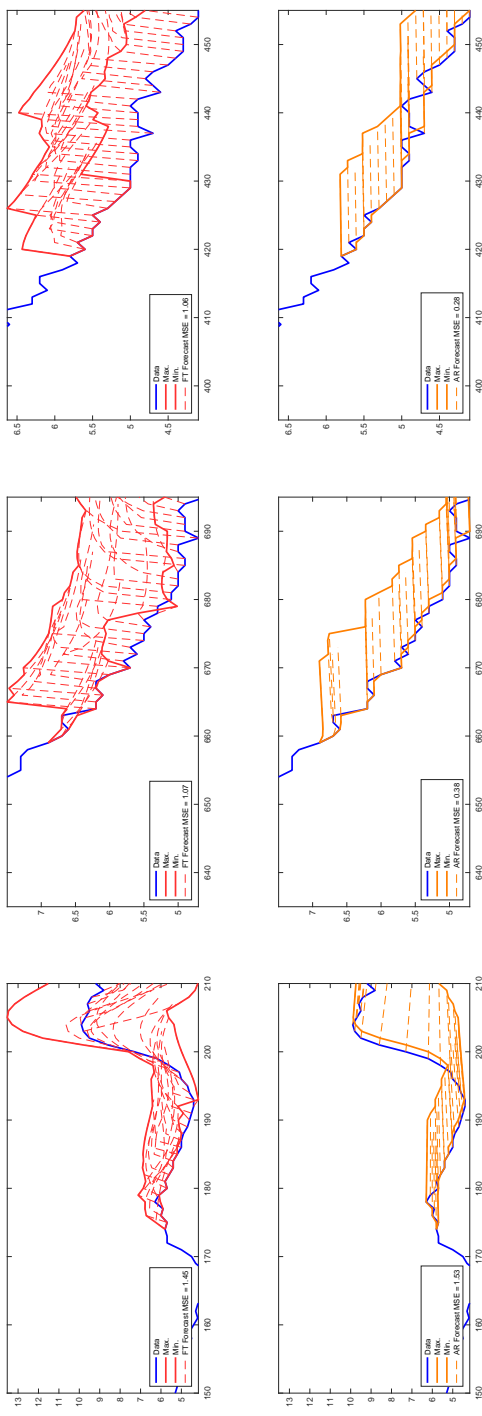
Figure 2: The 25 sets of 12-step ahead forecasts (red lines) for the US industrial production index (a) and US unemployment rate (b). In each panel, the forecasts for the SW (first column), NM (second column) and B (third column) datasets by applying FM (first line) and MAR (second line) models.
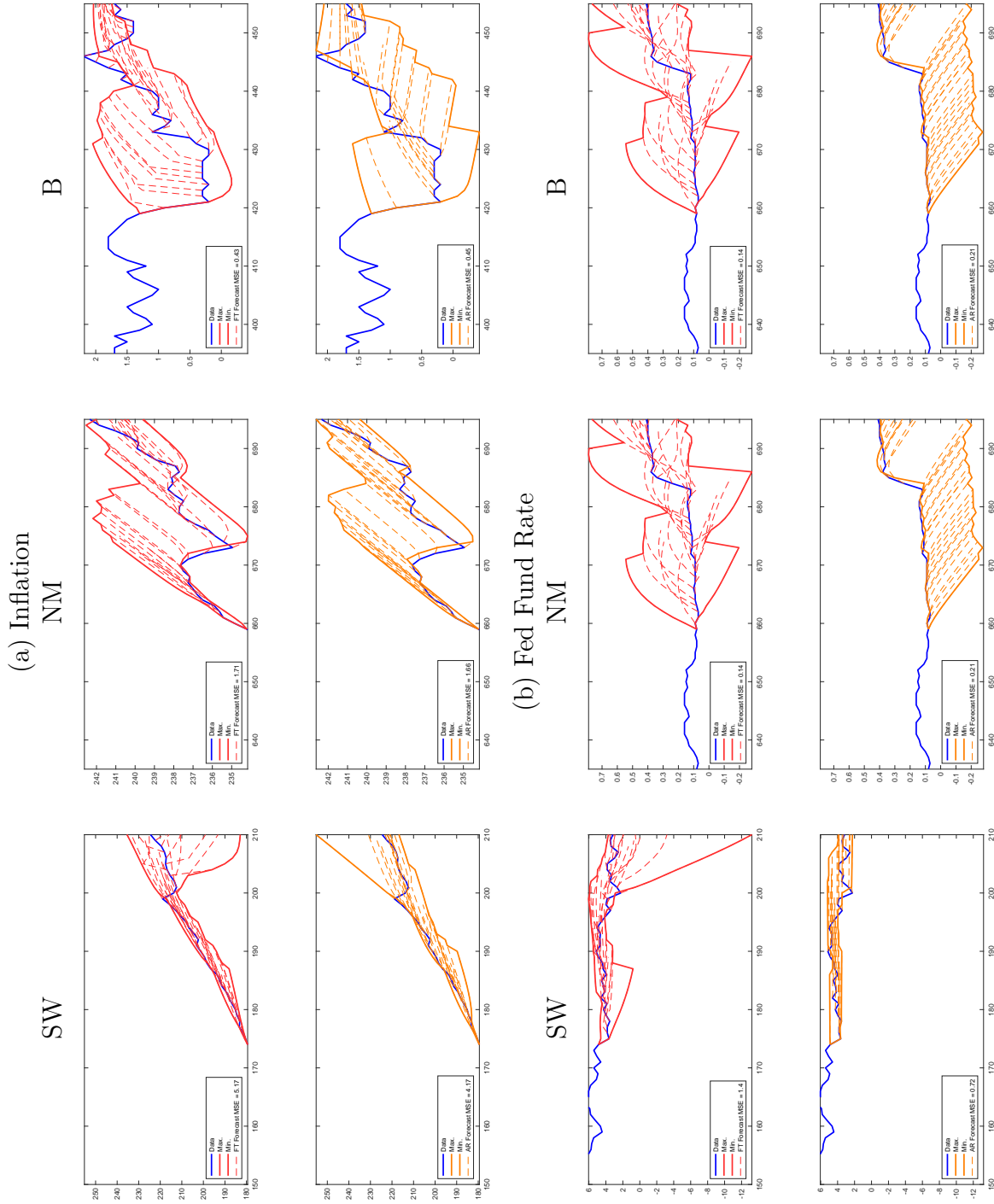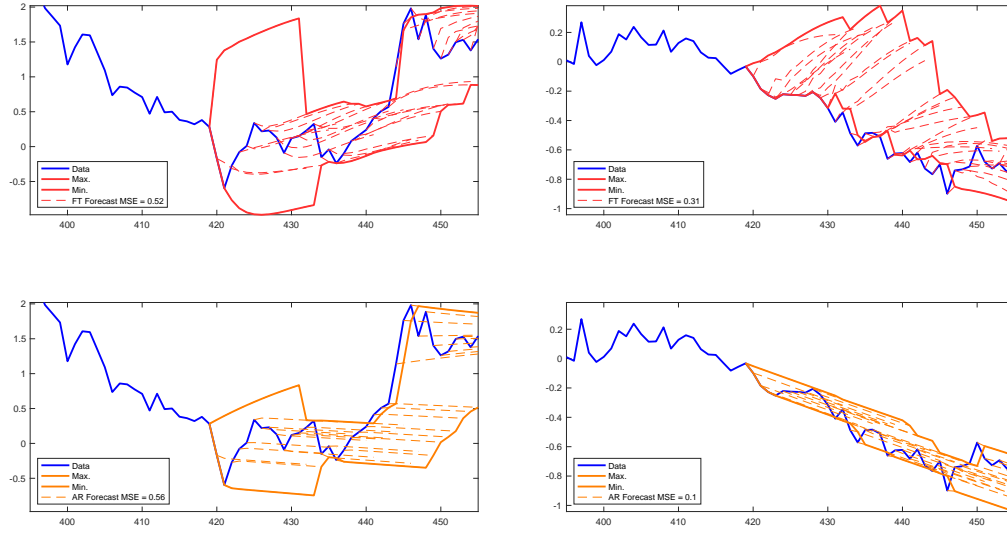
Figure 3: The 25 sets of 12-step ahead forecasts (red lines) for the US Inflation (a) and US Fed fund rate (b). In each panel, the forecasts for the SW (first column), NM (second column) and B (third column) datasets by applying FM (first line) and MAR (second line) models.
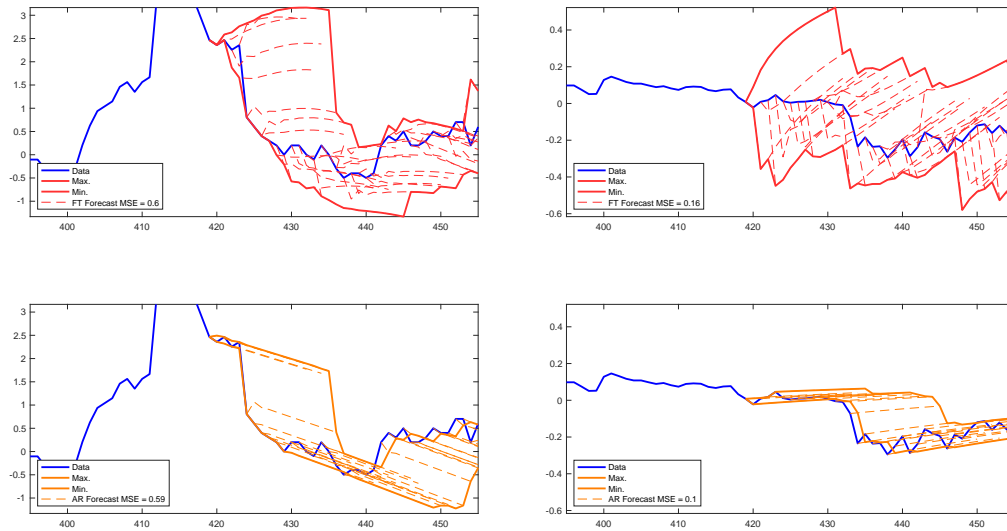
Figure 4: The 25 sets of 12-step-ahead forecasts (red lines) for the EU (a) and Japan (b) inflation (left) and interest rate (right) for the BL by applying FM (first line) and MAR (second line) models.
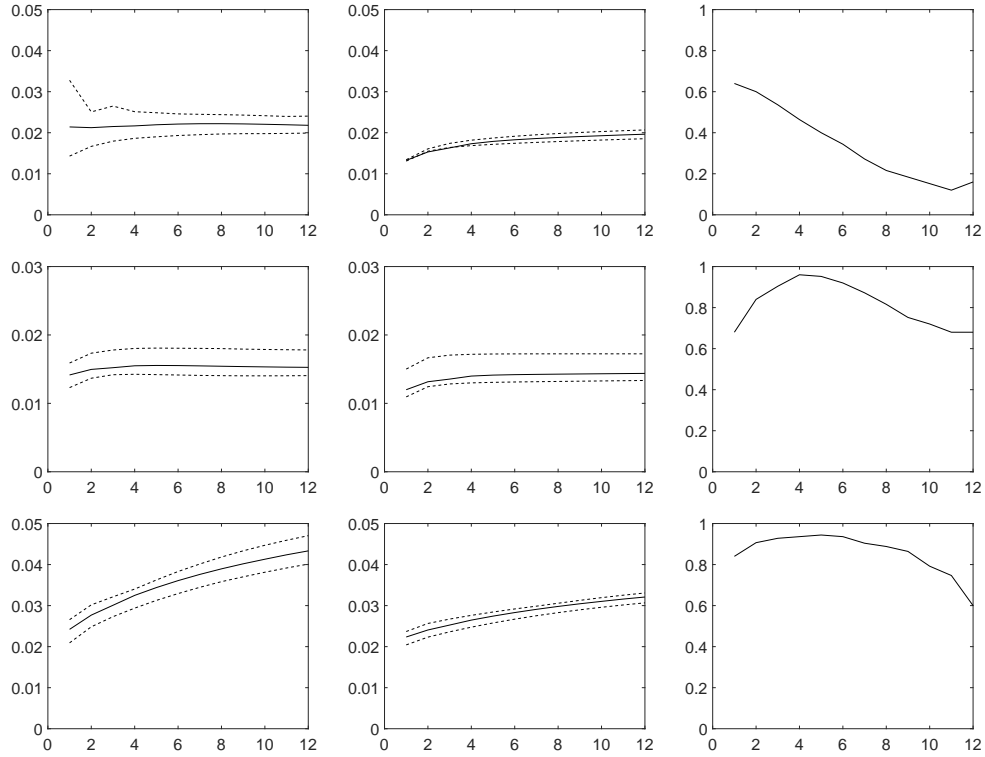
Figure 5: Estimates of the MSE in Th. 1-2 (solid line) and 5% and 95% quantiles, for the FM (first column) and MAR (second column) models. Ex-ante scoring rule performance (third column) measured as the frequency (vertical axis) of a true positive signal $f(s)$, over different horizons $s = 1, \ldots, 12$ (horizontal axis). Top: SW dataset. Middle: NM dataset. Bottom: BL dataset.
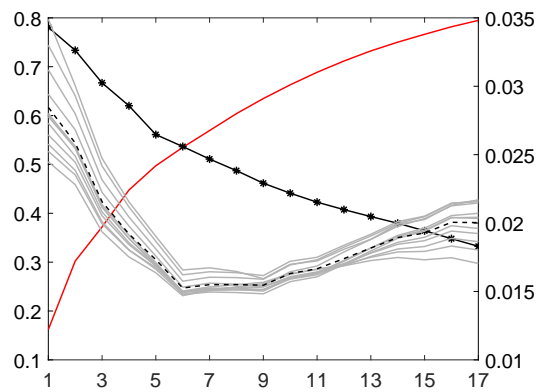
Figure 6: FM performance on the Bloomberg dataset when increasing the number of factors included. Proportion of variance (solid red line, left axis); adjusted R^2 (stars line, left axis) over number of factors considered (horizontal axis); average MSE (black line); and horizon-specific MSEs (gray lines).
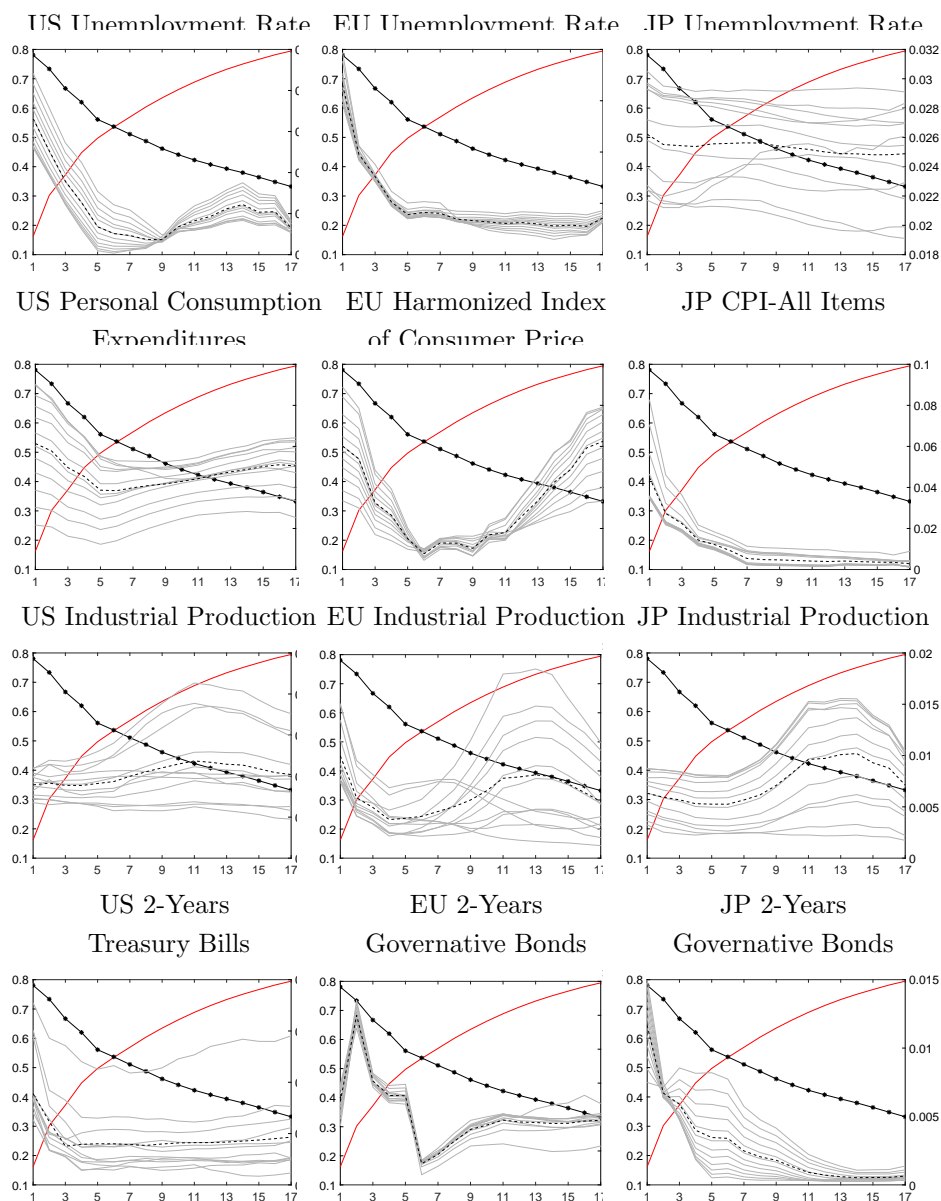
Figure 7: FM performance on the Bloomberg dataset when increasing the number of factors included. Proportion of variance (solid red line, left axis); adjusted R^2 (stars line, left axis) over number of factors considered (horizontal axis); average MSE (black line); and horizon-specific MSEs (gray lines).