Received: 11 March 2024 | Revised: 29 May 2024 | Accepted: 21 July 2024

DOI: 10.1002/for.3180

RESEARCH ARTICLE

WILEY

Forecasting Markov switching vector autoregressions: Evidence from simulation and application

Maddalena Cavicchioli

Department of Economics Marco Biagi, University of Modena and Reggio Emilia, Modena, Italy

Correspondence
Maddalena Cavicchioli, Department of Economics Marco Biagi, University of Modena and Reggio Emilia, Modena, Italy.
Email: maddalena.cavicchioli@unimore.it

Funding information
Work financially supported by a FAR 2022 research grant of the University of Modena and Reggio E., Italy

Abstract

We derive the optimal forecasts for multivariate autoregressive time series processes subject to Markov switching in regime. Optimality means that the trace of the mean square forecast error matrix is minimized by using suitable weighting observations. Then we provide neat analytic expressions for the optimal weights in terms of the matrices involved in a state space representation of the considered process. Our matrix expressions in closed form improve computational performance since they are readily programmable. Numerical simulations and an empirical application illustrate the feasibility of the proposed approach. We provide evidence that the forecasts using optimal weights increase forecast precision and are more accurate than the traditional Markov switching alternatives.

KEYWORDS

business cycle, forecasting, Markov switching vector autoregressive models, optimal weights, regime shifts, sample weighting observations, state space representation

1 | INTRODUCTION

Since the influential work of Hamilton (1989, 1990), Markov switching (MS) models have attracted considerable interest among econometricians to model various nonlinear observed time series in applied macroeconomics, which are subjected to change in regime. Flexibility is one of the main advantages of such models which become an appealing tool to capture the business cycle asymmetries, to investigate the cyclical behavior of many economic and financial variables arising from the real stock markets, and to detect "segmented trends" in macrovariables.

Stationarity, existence of moments, geometric ergodicity, statistical inference and asymptotic theory for MS vector autoregressive moving-average (MS VARMA) models, and their generalizations such as MS bilinear time series models, have been studied by several authors (see, e.g., Alvarez et al., 2017; Bibi & Ghezal, 2015; Cavicchioli, 2014a, 2014b, 2016; Douc et al., 2004; Francq & Zakoïan, 2001; Krolzig, 1997; and Stelzer, 2009).

In addition, spectral analysis of processes with time-varying coefficients is a relevant method for investigating both non-Gaussianity and nonlinearity, which continues to gain a growing interest. A formula in closed form for the spectral density matrix of MS VARMA models is given in Cavicchioli (2013). Matrix expressions for higher order moments and asymptotic Fisher information matrix of MS VARMA models are provided in Cavicchioli (2017a, 2017b), respectively. A recent spectral analysis approach for estimating the parameters involved in MS VAR models and for testing the fit of such models to the observed nonlinear data set has been proposed by Cavicchioli (2022). For a comprehensive survey on the literature on regime changes together with empirical applications in a number of areas of macroeconomics, see, for example, Hamilton (2016).

As remarked in Boot and Pick (2018), MS models have long been recognized to suffer from a discrepancy between in-sample and out-of-sample performances. More precisely, although MS models give a better in-sample fit with respect to linear models, they are usually outperformed by linear models in-out-sample forecasting; that is, in-sample, there are appealing results, while out-of-sample performances are frequently inferior to standard models. Examples in the literature are given to study the effect of uncertainty around states on forecasts. See, for example, Ang and Bekaert (2002), Clements and Krolzig (1998), Dacco and Satchell (1999), Krolzig (2000), Bessec and Bouabdallah (2005), Otranto (2005), and Nyberg (2018).

So, it is interesting to investigate the following question: *How forecast accuracy can be increased in* MS VAR *models*? In the univariate context, forecasts are typically evaluated by using mean square forecast error (MSFE) loss, which implies a bias-variance trade-off. Comparison of the forecast performances of MS and Threshold univariate AR models for US GNP can be found in Clements and Krolzig (1998). Methods for improving forecast accuracy in univariate MS models and a test to determine whether modeling a structural break improves forecasting have been provided by Boot and Pick, Boot and Pick (2018), (2020). Random subspace methods to forecast in high-dimensional regression settings can be found in Boot and Nibbering (2019). A new approach to forecasting time series that are subject to discrete structural breaks has been provided by Pesaran et al. (2006). These authors propose a Bayesian estimation and prediction procedure that allows for the possibility of new breaks occurring over the forecast horizon, taking account of the size and duration of past breaks (if any) by means of a hierarchiral hidden Markov chain model.

Given the focus on out-of-sample forecast, we include some relevant references, which are related to our research topic. Clements et al. (2004) discuss the current state-of-the-art in estimating, evaluating, and selecting among nonlinear forecasting models for economic and financial time series. Elliott and Timmermann (2005) propose a new forecast combination method that lets the combination weights be driven by regime switching in a latent state variable and show that such a method performs better than a range of alternative combination schemes for a variety of macroeconomic variables. A flexible approach to combine forecasts of future spot rates with forecasts from time series models or macroeconomic variables has been developed by Guidolin and Timmermann (2009). These authors find empirical evidence that, accounting for both regimes in interest rate dynamics and combining forecasts from different models, helps improve the out-of-sample forecasting performance for US short-term rates. Hou (2017) investigates the forecast performances of infinite hidden MS VARs for some financial and macroeconomic applications. Kundu and Paul (2022) examine the effect of economic policy uncertainty (EPU) on stock market return and volatility under heterogeneous market characteristics. Their estimation results suggest that the impact of EPU is significant in the bear market, and it is negligible in the bull market.

Empirical applications include forecasting interest rates and US business cycle via MS VAR (Kontolemis, 2001; Nyberg, 2018), volatility forecasting with double MS GARCH (Chen et al., 2009), forecasting exchange rates via MS models (Parikakis & Merika, 2009; Nikolsko-Rzhevskyy & Prodan, 2012), prediction of GDP growth and business cycle turning points in the Euro area via MS mixed-frequency VAR (Foroni et al., 2015), forecasting risk with MS GARCH (Alemohammad et al., 2020; Ardia et al., 2018), and forecasting US inflation using Markov dimension switching (Prüser, 2021).

The purpose of our paper is to give a contribution to the existing literature on forecasting for MS models. More precisely, the aim is to circumvent the above-mentioned problems about forecast accuracy presenting a nice matrix machinery which produces optimal forecasts for the class of MS VAR(p) processes. So, we contribute to the existing literature in threefolds. First, we derive optimal forecasts for MS VAR(p) models by minimizing the quadratic function given by the trace of the expected MSFE matrix by using sample weighting observations. Then we provide explicit neat matrix expressions in closed form for the forecasting optimal weights in terms of the matrices involved in a state space representation of the model specification. These formulas improve computational performance since they are readily programmable. Our results generalize to the most general setting the work of Boot and Pick (2018) regarding univariate MS AR(0) models. Second, our matrix analysis, which is different to that employed up to now in the literature, is based on an indirect construction via a Markovian representation of the considered MS VAR model. This simplifies the computation, produces simple and easily tractable matrix expressions, and provides a unified framework for forecasting different types of MS vector autoregressive processes. The main advantage of the proposed approach is related to the simplicity of the derived mathematical expressions that eases the computational effort also in the general case of multivariate autoregressive models with regime-dependent parameters. Third, theoretical results and methods are useful tools for practitioners, providing evidence with numerical and empirical applications that forecasts with optimal weights improves the forecast accuracy. Performances of the optimal weights through simulations and an empirical application confirm the issues of Boot and Pick (2018) also in the multivariate case with regime-dependent autoregressive coefficients. That is, the forecasts using optimal weights increase forecast precision and are more accurate than linear alternatives.

The paper is structured as follows. Section 2 lays out the basic model, assumptions and notation. A state space representation of a VAR(0) process with Markovian regime shifts in the intercept term and in the variance is derived. The algorithm for the extraction of optimal weights is applied, and the issues on forecasting are discussed. Section 3 illustrates the results for the general case of a MS VAR(p) model with regime dependent parameters. Numerical simulations and an empirical application to interest rate and term spread in US area are discussed in Section 4. Section 5 concludes the study. See Hamilton ((1994), §4 and §22) and Krolzig (1997) as basic references on MS autoregressive processes.

2 | FORECASTING TIME SERIES PROCESSES WITH MS

In this section, we first consider dynamic vector models where the intercept term and the variance-covariance matrix of the innovation are subject to occasional discrete shifts, and postpone the general case with time-varying autoregressive part.

2.1 | The basic model

Let $\boldsymbol{y}_t$ be a K-dimensional random vector with values in $\mathbb{R}^K$. Suppose that $\boldsymbol{y}_t$ is driven by the following MS VAR(0) model with a M-state MS intercept and variance, in short MS(M) VAR(0):

$$\boldsymbol{y}_t = \boldsymbol{\nu}_{s_t} + \boldsymbol{\Sigma}_{s_t} \boldsymbol{u}_t \tag{1}$$

where the intercept $\boldsymbol{\nu}_{s_t}$ is $K \times 1$, $\boldsymbol{\Sigma}_{s_t}$ is a $K \times K$ nonsingular matrix, and $\boldsymbol{u}_t \sim$ i.i.d. $(\boldsymbol{0}, \boldsymbol{I}_K)$. Here $\boldsymbol{I}_K$ denotes the $K \times K$ identity matrix. The variance–covariance matrix $\boldsymbol{\Omega}_{s_t} = \boldsymbol{\Sigma}_{s_t}\boldsymbol{\Sigma}'_{s_t} \in \mathbb{R}^{K \times K}$ is positive definite. As usual, $\mathbb{R}^{m \times n}$ denotes the class of real $m \times n$ matrices and $\mathbb{R}^n$ the class of n-dimensional real vectors, that is, $\mathbb{R}^n = \mathbb{R}^{n \times 1}$. Such MS models serve well to depict the business cycle of various industrialized economies as shown, for example, in Knuppel (2009).

Assumption 1. The process (s_t) in (1) is an irreducible and aperiodic (hence ergodic) Markov chain with values in the set $\Xi = \{1, \ldots, M\}$, stationary transition probabilities denoted by $p_{ij} = Pr(s_t = j | s_{t-1} = i)$, for $i, j = 1, \ldots, M$, and unconditional (or steady state) probabilities $\pi_i = Pr(s_t = i)$, for $i = 1, \ldots, M$.

Collect p_{ij} and π_i into a $M \times M$ matrix $\boldsymbol{P} = (p_{ij})$ and a $M \times 1$ vector $\boldsymbol{\pi} = (\pi_1 \ldots \pi_M)'$, called the *transition probability matrix* and the *stationary vector* of the chain, respectively. *Ergodicity* implies the existence of $\boldsymbol{\pi}$ satisfying $\boldsymbol{\pi} = \boldsymbol{P}'\boldsymbol{\pi}$ and $\boldsymbol{i}'_M \boldsymbol{\pi} = 1$, where $\boldsymbol{i}_M$ denotes the $M \times 1$ vector of ones. *Irreducibility* implies that $\pi_i > 0$, for $i = 1, \ldots, M$, meaning that all unobservable states are possible. It is also assumed that $\pi_1 \geq \pi_2 \geq \ldots \geq \pi_M$ for identifiability of regimes, and $\boldsymbol{\nu}_i \neq \boldsymbol{\nu}_j$ for $i, j = 1, \ldots, M$, $i \neq j$. In order to clarify the definition of aperiodic Markov chain, we need some preliminaries. The period d_i of a state $i \in \Xi$ is given by $d_i := \gcd\{m \geq 1 : p_{ii}^{(m)} > 0\}$, where gcd denotes the greatest common divisor and $p_{ii}^{(m)} = Pr(s_{t+m} = i | s_t = i)$. Set $d_i = \infty$ if $p_{ii}^{(m)} = 0$ for all $m \geq 1$. If $d_i = 1$, then the state $i \in \Xi$ is called aperiodic. A Markov chain is said to be *aperiodic* if all its states are aperiodic.

Assumption 2. The pair $(s_t, \boldsymbol{u}_t)$ is a strictly stationary process defined in some probability space, and the shock $(\boldsymbol{u}_t)$ is independent of the Markov chain (s_t).

In addition, we assume that the process $(\boldsymbol{y}_t)$ is second-order stationary; that is, it satisfies the matrix condition in Theorem 2 from Francq and Zakoïan (2001).

2.2 | State space representation

Following Hamilton ((1994), §22) and Krolzig ((1997), §2), a useful representation for the Markov chain (s_t) is obtained by letting $\boldsymbol{\xi}_t$ denote the $M \times 1$ vector whose ith element is equal to unity if $s_t = i$ and zero otherwise. Then the Markov chain follows a VAR(1) process

$$\boldsymbol{\xi}_t = \boldsymbol{P}'\boldsymbol{\xi}_{t-1} + \boldsymbol{v}_t \tag{2}$$

where $\boldsymbol{v}_t = \boldsymbol{\xi}_t - E(\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t-1})$ is a zero mean martingale difference sequence.

Set $\boldsymbol{D} = \text{diag}(\pi_1 \ldots \pi_M) \in \mathbb{R}^{M\times M}$. Then we have $E(\boldsymbol{\xi}_t) = \boldsymbol{\pi}$ and $E(\boldsymbol{\xi}_t\boldsymbol{\xi}'_{t+h}) = \boldsymbol{DP}^h$ for every $h \geq 0$. In particular, $E(\boldsymbol{\xi}_t\boldsymbol{\xi}'_t) = \boldsymbol{D}$. It follows that $\text{var}(\boldsymbol{\xi}_t) = \boldsymbol{D} - \boldsymbol{\pi\pi}'$ is singular due to the adding-up restriction $\boldsymbol{i}'_M\boldsymbol{\pi} = 1$. Furthermore, $\boldsymbol{v}_t \sim$ i.i.d. $(\boldsymbol{0}, \boldsymbol{\Sigma_v})$, where $\boldsymbol{\Sigma_v} = \boldsymbol{D} - \boldsymbol{P}'\boldsymbol{DP}$.

Let $\boldsymbol{\nu}_i$ and $\boldsymbol{\Sigma}_i$ be obtained from $\boldsymbol{\nu}_{s_t}$ and $\boldsymbol{\Sigma}_{s_t}$, respectively, by setting $s_t = i$, for $i = 1, \ldots, M$. Define

$$\boldsymbol{\Lambda} = (\boldsymbol{\nu}_1 \ldots \boldsymbol{\nu}_M) \in \mathbb{R}^{K\times M} \qquad \boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_1 \ldots \boldsymbol{\Sigma}_M) \in \mathbb{R}^{K\times(KM)}.$$

Then we get the state space representation of (1):

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{\Lambda\xi}_t + \boldsymbol{\Sigma}(\boldsymbol{\xi}_t \otimes \boldsymbol{I}_K)\boldsymbol{u}_t \\ \boldsymbol{\xi}_t &= \boldsymbol{P}'\boldsymbol{\xi}_{t-1} + \boldsymbol{v}_t. \end{aligned} \tag{3}$$

The ML (or OLS) estimates of the population parameters $\boldsymbol{\nu}_i$ and $\boldsymbol{\Omega}_i = \boldsymbol{\Sigma}_i\boldsymbol{\Sigma}'_i$ in model (1), or (3), are given by

$$\hat{\boldsymbol{\nu}}_i = \left[\sum_{t=1}^{T} \hat{\xi}_{it|T}\right]^{-1} \left[\sum_{t=1}^{T} \hat{\xi}_{it|T}\boldsymbol{y}_t\right] \tag{4}$$

and

$$\hat{\boldsymbol{\Omega}}_i = \left[\sum_{t=1}^{T} \hat{\xi}_{it|T}\right]^{-1} \left[\sum_{t=1}^{T} \hat{\xi}_{it|T}(\boldsymbol{y}_t - \hat{\boldsymbol{\nu}}_i)(\boldsymbol{y}_t - \hat{\boldsymbol{\nu}}_i)'\right] \tag{5}$$

for $i = 1, \ldots, M$. See, for example, Cavicchioli (2014b, 2021). A fast algorithm for determining the smoothed regime probabilities $\hat{\boldsymbol{\xi}}_{t|T} = E[\boldsymbol{\xi}_t|\boldsymbol{Y}_T]$, for $t = 1, \ldots, T$, can be found in Hamilton ((1994), §22) and Krolzig ((1997), §5). Here $\boldsymbol{Y}_T$ denotes the information set that is available up to time T, that is, $\boldsymbol{Y}_T = \{\boldsymbol{y}_T, \boldsymbol{y}_{T-1}, \ldots\}$, and ξ_{it} (resp. $\hat{\xi}_{it|T}$) is the ith component of the $M \times 1$ vector $\boldsymbol{\xi}_t$ (resp. $\hat{\boldsymbol{\xi}}_{t|T}$), for $i = 1, \ldots, M$.

2.3 | MS forecast

The MS forecast for model (1) is given by the conditional expectation

$$\hat{\boldsymbol{y}}_{T+1|T} = E[\boldsymbol{y}_{T+1}|\boldsymbol{Y}_T] \tag{6}$$

from the state space representation (3), that is,

$$\hat{\boldsymbol{y}}_{T+1|T} = \hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\xi}}_{T+1|T} = \sum_{i=1}^{M} \hat{\xi}_{i,T+1|T}\hat{\boldsymbol{\nu}}_i \tag{7}$$

where $\hat{\boldsymbol{\xi}}_{T+1|T} = E[\boldsymbol{\xi}_{T+1}|\boldsymbol{Y}_T] = \boldsymbol{P}'\hat{\boldsymbol{\xi}}_{T|T}$ is the vector of the predicted regime probabilities in the forecast period. The i-th component of $\boldsymbol{\xi}_{T+1|T}$ is $\xi_{i,T+1|T} = \sum_{j=1}^{T} p_{ji}\xi_{jT|T}$.

Substituting (4) into (7) yields

$$\hat{\boldsymbol{y}}_{T+1|T} = \sum_{t=1}^{T} w_{MS,t}\boldsymbol{y}_t \tag{8}$$

where the standard *MS weights* are given by

$$w_{MS,t} = \sum_{i=1}^{M} \left[\sum_{t=1}^{T} \hat{\xi}_{it|T}\right]^{-1} \hat{\xi}_{i,T+1|T}\,\hat{\xi}_{it|T}. \tag{9}$$

Collect the MS weights in a $T \times 1$ vector $\boldsymbol{w}_{MS} = (w_{MS,1}, \ldots, w_{MS,T})'$. It is easily verified that $\boldsymbol{i}'_T\boldsymbol{w}_{MS} = 1$; that is, the MS weights sum to one. Here $\boldsymbol{i}_T$ denotes the $T \times 1$ vector of ones.

Set $\boldsymbol{e}_{T+1} = \boldsymbol{y}_{T+1} - \hat{\boldsymbol{y}}_{T+1|T}$. The forecast error $\boldsymbol{e}_{T+1}$ can be decomposed according to three causes of uncertainty as follows. The first relates with the term $\boldsymbol{\Lambda P}'(\xi_T - \hat{\boldsymbol{\xi}}_{T|T})$, and it is caused by the regime classification errors (also called filter uncertainty). The summands $\boldsymbol{\Lambda v}_{T+1}$ and $\boldsymbol{\Sigma}(\boldsymbol{\xi}_{T+1} \bigotimes \boldsymbol{I}_K)\boldsymbol{u}_{T+1}$ reflect the uncertainty due to future system shocks $\boldsymbol{v}_{T+1}$ and $\boldsymbol{u}_{T+1}$.

Then the expected MSFE matrix is given by

$$\begin{aligned} E\left[\boldsymbol{e}_{T+1}\boldsymbol{e}'_{T+1}|\boldsymbol{Y}_T\right] = \; & \hat{\boldsymbol{\Lambda}}E\left[\left(\boldsymbol{\xi}_{T+1} - \hat{\boldsymbol{\xi}}_{T+1|T}\right)\right. \\ & \left.\left(\boldsymbol{\xi}_{T+1} - \hat{\boldsymbol{\xi}}_{T+1|T}\right)'|\boldsymbol{Y}_T\right]\hat{\boldsymbol{\Lambda}}' \\ & + \hat{\boldsymbol{\Sigma}}\left(E\left[\boldsymbol{\xi}_{T+1}\boldsymbol{\xi}'_{T+1}|\boldsymbol{Y}_T\right] \otimes \boldsymbol{I}_K\right)\hat{\boldsymbol{\Sigma}}' \end{aligned} \tag{10}$$

where

$$E\left[\boldsymbol{\xi}_{T+1}\boldsymbol{\xi}'_{T+1}|\boldsymbol{Y}_T\right] = \text{diag}(\hat{\boldsymbol{\xi}}_{T+1|T}) = \text{diag}(\boldsymbol{P}'\hat{\boldsymbol{\xi}}_{T|T})$$

and

$$\begin{aligned} E\left[\left(\boldsymbol{\xi}_{T+1} - \hat{\boldsymbol{\xi}}_{T+1|T}\right)\left(\boldsymbol{\xi}_{T+1} - \hat{\boldsymbol{\xi}}_{T+1|T}\right)'|\boldsymbol{Y}_T\right] & \\ = \text{diag}(\hat{\boldsymbol{\xi}}_{T+1|T}) - \hat{\boldsymbol{\xi}}_{T+1|T}\hat{\boldsymbol{\xi}}'_{T+1|T} & \\ = \text{diag}(\boldsymbol{P}'\hat{\boldsymbol{\xi}}_{T|T}) - \boldsymbol{P}'\hat{\boldsymbol{\xi}}_{T|T}\hat{\boldsymbol{\xi}}'_{T|T}\boldsymbol{P}. & \end{aligned}$$

2.4 | Optimal forecast

Following the nice work of Boot and Pick (2018) in the univariate setting, we derive the optimal forecast by using a weighted average of the observations with weights that minimize the quadratic function defined by the trace of the MSFE matrix. More precisely, the forecast from weighted observations is defined as

$$\hat{\boldsymbol{y}}_{\boldsymbol{w},T+1} = \sum_{t=1}^{T} w_t \boldsymbol{y}_t = (\boldsymbol{w}' \otimes \boldsymbol{I}_K)\boldsymbol{y} \tag{11}$$

where $\boldsymbol{w} = (w_1, ..., w_T)'$ and $\boldsymbol{y} = (\boldsymbol{y}_1', ..., \boldsymbol{y}_T')'$.

We discuss two minimization problems. First, we do not restrict the weights to be positive and to sum to one as satisfied for MS weights $\boldsymbol{w}_{MS}$ above. Call them *unrestricted optimal* (UO) weights, which are denoted by $\boldsymbol{w}_{UO}$. Second, we impose that the weights (not necessarily positive) satisfy the restriction $\boldsymbol{i}_T'\boldsymbol{w} = 1$ (as in the quoted paper) since an identifying condition could be required. Call them *restricted optimal* (RO) weights, which are denoted by $\boldsymbol{w}_{RO}$. In all cases for the considered MS models, we compare the square root of the trace (that is, the module) of the obtained expected MSFE matrices.

So, we choose the optimal forecast $\hat{\boldsymbol{y}}_{\boldsymbol{w},T+1}$ for the considered MS models so as to minimize the quadratic function defined by the trace (tr) of the MSFE matrix $E\left[\boldsymbol{e}_{\boldsymbol{w},T+1}\boldsymbol{e}_{\boldsymbol{w},T+1}'\right]$, where $\boldsymbol{e}_{\boldsymbol{w},T+1} = \boldsymbol{y}_{T+1} - \hat{\boldsymbol{y}}_{\boldsymbol{w},T+1}$ and $\hat{\boldsymbol{y}}_{\boldsymbol{w},T+1}$ is as in (11).

Ignoring the parameter estimation problem, the MSFE matrix for the MS model (1) is given in vectorial form by

$$\begin{aligned}\text{MSFE}(\boldsymbol{w}) &= E\left[\boldsymbol{e}_{\boldsymbol{w},T+1}\boldsymbol{e}_{\boldsymbol{w},T+1}'\right] = \boldsymbol{\Lambda}\boldsymbol{D}\boldsymbol{\Lambda}' \\ &-\boldsymbol{\Lambda}E[\boldsymbol{\xi}_{T+1}\boldsymbol{s}'](\boldsymbol{\Lambda}'\otimes\boldsymbol{I}_T)(\boldsymbol{w}\otimes\boldsymbol{I}_K) \\ &-(\boldsymbol{w}'\otimes\boldsymbol{I}_K)(\boldsymbol{\Lambda}\otimes\boldsymbol{I}_T)E[\boldsymbol{s}\boldsymbol{\xi}_{T+1}']\boldsymbol{\Lambda}' \\ &+(\boldsymbol{w}'\otimes\boldsymbol{I}_K)(\boldsymbol{\Lambda}\otimes\boldsymbol{I}_T)E[\boldsymbol{s}\boldsymbol{s}'](\boldsymbol{\Lambda}'\otimes\boldsymbol{I}_T) \\ &(\boldsymbol{w}\otimes\boldsymbol{I}_K) + \boldsymbol{\Sigma}(\boldsymbol{D}\otimes\boldsymbol{I}_K)\boldsymbol{\Sigma}' \\ &+(\boldsymbol{w}'\otimes\boldsymbol{I}_K)(\boldsymbol{\Sigma}\otimes\boldsymbol{I}_T)(\boldsymbol{D}\otimes\boldsymbol{I}_{TK}) \\ &(\boldsymbol{\Sigma}'\otimes\boldsymbol{I}_T)(\boldsymbol{w}\otimes\boldsymbol{I}_K)\end{aligned} \tag{12}$$

where $\boldsymbol{s} = (\boldsymbol{\xi}_1', ..., \boldsymbol{\xi}_T')' \in \mathbb{R}^{TM}$.

The first derivatives of the trace of MSFE($\boldsymbol{w}$) with respect to $\boldsymbol{w} \otimes \boldsymbol{I}_K$ are the $(TK) \times K$ matrix given by

$$\frac{\partial\, \text{trMSFE}(\boldsymbol{w})}{\partial \boldsymbol{w}\otimes\boldsymbol{I}_K} = 2\boldsymbol{M}(\boldsymbol{w}\otimes\boldsymbol{I}_K) - 2\boldsymbol{N} \tag{13}$$

where

$$\begin{aligned}\boldsymbol{M} &= (\boldsymbol{\Sigma}\otimes\boldsymbol{I}_T)(\boldsymbol{D}\otimes\boldsymbol{I}_{TK})(\boldsymbol{\Sigma}'\otimes\boldsymbol{I}_T) \\ &+(\boldsymbol{\Lambda}\otimes\boldsymbol{I}_T)E[\boldsymbol{s}\boldsymbol{s}'](\boldsymbol{\Lambda}'\otimes\boldsymbol{I}_T)\end{aligned} \tag{14}$$

and

$$\boldsymbol{N} = (\boldsymbol{\Lambda}\otimes\boldsymbol{I}_T)E[\boldsymbol{s}\boldsymbol{\xi}_{T+1}']\boldsymbol{\Lambda}'. \tag{15}$$

Vanishing the first derivatives in (13) yields the following result:

Theorem 1. With the above notation, the unrestricted optimal weights for the $MS(M)$ $VAR(0)$ in (1) with Markovian representation (3) are given by

$$\boldsymbol{w}_{UO}\otimes\boldsymbol{I}_K = \boldsymbol{M}^{-1}\boldsymbol{N} \tag{16}$$

where $\boldsymbol{M}$ and $\boldsymbol{N}$ are as in (14) and (15), respectively.

The $(TK)\times(TK)$ matrix $\boldsymbol{M}$ is invertible. To see this, we factorize $\boldsymbol{M} = \boldsymbol{M}_1 + \boldsymbol{M}_2$. Now the first matrix summand $\boldsymbol{M}_1 = [\boldsymbol{\Sigma}(\boldsymbol{D} \otimes \boldsymbol{I}_K)\boldsymbol{\Sigma}'] \otimes \boldsymbol{I}_T$ is positive definite as $\boldsymbol{\Sigma}$ has full rank K and $\boldsymbol{D}$ is a diagonal matrix with positive entries $\pi_i > 0$, for $i = 1, ..., M$. The second matrix summand $\boldsymbol{M}_2 = (\boldsymbol{\Lambda} \otimes \boldsymbol{I}_T)E[\boldsymbol{s}\boldsymbol{s}'](\boldsymbol{\Lambda}' \otimes \boldsymbol{I}_T)$ is positive semidefinite. Thus, $\boldsymbol{M}$ is positive definite; hence, it is invertible.

The unrestricted optimal weights in (16) really minimize the quadratic function trMSFE($\boldsymbol{w}$) as the second derivatives

$$\frac{\partial^2 \text{trMSFE}(\boldsymbol{w})}{\partial \boldsymbol{w} \otimes \boldsymbol{I}_K\, \partial \boldsymbol{w}' \otimes \boldsymbol{I}_K} = 2\boldsymbol{M} \tag{17}$$

is positive definite.

Substituting (16) into (11) and (12) gives the (unrestricted) optimal forecast and the minimum (in the sense of matrix norm) MSFE, respectively. Furthermore, to make our matrix expressions suitable for computation, we calculate the expectations in (12) and (16) conditional on $\boldsymbol{Y}_T$ under the assumption that the states are uncertain and use the ML estimates of the population parameters.

Minimizing the function tr$E[\boldsymbol{e}_{\boldsymbol{w},T+1}\boldsymbol{e}_{\boldsymbol{w},T+1}']$, subject to $(\boldsymbol{w}' \otimes \boldsymbol{I}_K)\boldsymbol{y} = 1$, gives the following result.

Theorem 2. Let us consider the $MS(M)$ $VAR(0)$ model in (1) with a Markovian representation (3). Then the restricted optimal weights (RO) are given by

$$\boldsymbol{w}_{RO}\otimes\boldsymbol{I}_K=\boldsymbol{M}^{-1}\boldsymbol{N}+\frac{K-b}{a}\boldsymbol{M}^{-1}(\boldsymbol{i}_T\otimes\boldsymbol{I}_K) \tag{18}$$

where

$$\begin{aligned} a &= \mathrm{tr}\left[(\boldsymbol{i}_T'\otimes\boldsymbol{I}_K)\boldsymbol{M}^{-1}(\boldsymbol{i}_T\otimes\boldsymbol{I}_K)\right] \\ b &= \mathrm{tr}\left[(\boldsymbol{i}_T'\otimes\boldsymbol{I}_K)\boldsymbol{M}^{-1}\boldsymbol{N}\right] \end{aligned}$$

and $\boldsymbol{M}$ and $\boldsymbol{N}$ are as in (14) and (15), respectively.

Substituting (18) into (11) and (12) gives the (restricted) optimal forecast and the minimum (in the sense of matrix norm) MSFE, respectively.

For practical inference purposes, we adopt a plug-in approach where the involved matrices are replaced by their maximum likelihood (ML) estimates.

3 | FORECASTING GENERAL AUTOREGRESSIVE VECTOR PROCESSES WITH MS

Let us consider a M-state Markov switching VAR(p) model, in short MS(M) VAR(p), of the following type

$$\boldsymbol{y}_t=\boldsymbol{\nu}_{s_t}+\sum_{i=1}^{p}\Phi_{i,s_t}\boldsymbol{y}_{t-i}+\boldsymbol{\Sigma}_{s_t}\boldsymbol{u}_t \tag{19}$$

where $\boldsymbol{y}_t$ is a K-dimensional random vector with values in $\mathbb{R}^K$, the autoregressive coefficients are regime dependent $(K\times K)$ matrices, and the assumptions in Section 2 hold.

Following Krolzig ((1997), §2), model (19) can be written as

$$\boldsymbol{y}_t=(\boldsymbol{x}_t'\otimes\boldsymbol{I}_K)\boldsymbol{\beta}_{s_t}+\boldsymbol{\Sigma}_{s_t}\boldsymbol{u}_t \tag{20}$$

where

$$\boldsymbol{x}_t=(1\,\boldsymbol{y}_{t-1}'\ldots\boldsymbol{y}_{t-p}')'\in\mathbb{R}^R$$

and

$$\boldsymbol{\beta}_{s_t}=\left(\boldsymbol{\nu}_{s_t}'\,[\mathrm{vec}\boldsymbol{\Phi}_{1,s_t}]'\ldots[\mathrm{vec}\boldsymbol{\Phi}_{p,s_t}]'\right)'\in\mathbb{R}^{RK}$$

with $R=pK+1$. As before, let $\boldsymbol{\beta}_i$ and $\boldsymbol{\Omega}_i$ be obtained from $\boldsymbol{\beta}_{s_t}$ and $\boldsymbol{\Omega}_{s_t}=\boldsymbol{\Sigma}_{s_t}\boldsymbol{\Sigma}_{s_t}'$, respectively, by setting $s_t=i$, for $i=1,\ldots,M$.

The ML (or OLS) estimates of the population parameters $\boldsymbol{\beta}_i$ and $\boldsymbol{\Omega}_i$ for the MS(M) VAR(p) model in (20) are given by

$$\hat{\boldsymbol{\beta}}_i=\left[\boldsymbol{X}_{iT}^{-1}\otimes\boldsymbol{I}_K\right]\left[\sum_{t=1}^{T}\hat{\xi}_{it|T}(\boldsymbol{x}_t\otimes\boldsymbol{I}_K)\boldsymbol{y}_t\right] \tag{21}$$

and

$$\hat{\boldsymbol{\Omega}}_i=\left[\sum_{t=1}^{T}\hat{\xi}_{it|T}\right]^{-1}\left[\sum_{t=1}^{T}\hat{\xi}_{it|T}\hat{\epsilon}_{t,i}\hat{\epsilon}_{t,i}'\right] \tag{22}$$

where

$$\hat{\epsilon}_{t,i}=\boldsymbol{y}_t-(\boldsymbol{x}_t'\otimes\boldsymbol{I}_K)\hat{\boldsymbol{\beta}}_i$$

for $i=1,\ldots,M$. See, for example, Cavicchioli (2014b, 2021). Here the assumption is that the $(R\times R)$ matrix

$$\boldsymbol{X}_{iT}=\sum_{t=1}^{T}\hat{\xi}_{it|T}\boldsymbol{x}_t\boldsymbol{x}_t'$$

is invertible for every $i=1,\ldots,M$. This extends the analogous one (case $M=1$) for linear VAR(p) models. See Hamilton ((1994), §11).

Model (20) has the following state space representation

$$\begin{aligned} \boldsymbol{y}_t &= (\boldsymbol{x}_t'\otimes\boldsymbol{I}_K)\boldsymbol{B}\boldsymbol{\xi}_t+\boldsymbol{\Sigma}(\boldsymbol{\xi}_t\otimes\boldsymbol{I}_K)\boldsymbol{u}_t \\ \boldsymbol{\xi}_t &= \boldsymbol{P}'\boldsymbol{\xi}_{t-1}+\boldsymbol{v}_t \end{aligned} \tag{23}$$

where $\boldsymbol{B}=(\boldsymbol{\beta}_1\ldots\boldsymbol{\beta}_M)\in\mathbb{R}^{(RK)\times M}$ and $\boldsymbol{\Sigma}=(\boldsymbol{\Sigma}_1\ldots\boldsymbol{\Sigma}_M)\in\mathbb{R}^{K\times(KM)}$.

Since $\boldsymbol{x}_t'\otimes\boldsymbol{I}_K$ is deterministic given $\boldsymbol{Y}_T$, the expectation of $\boldsymbol{y}_{T+1}$ conditional $\boldsymbol{Y}_T$ is given by

$$\hat{\boldsymbol{y}}_{T+1|T}=(\boldsymbol{x}_{T+1}'\otimes\boldsymbol{I}_K)\hat{\boldsymbol{B}}\hat{\boldsymbol{\xi}}_{T+1|T}=\sum_{i=1}^{M}(\boldsymbol{x}_{T+1}'\otimes\boldsymbol{I}_K)\hat{\boldsymbol{\beta}}_i\hat{\xi}_{i,T+1|T}. \tag{24}$$

Substituting (21) into (24) and doing standard matrix computation yield formula (8) where the MS weights now are

$$w_{MS,t} = \sum_{i=1}^{M} \hat{\xi}_{i,T+1|T}\hat{\xi}_{i,t|T}\left(\boldsymbol{x}_{T+1}^{'}\boldsymbol{X}_{iT}^{-1}\boldsymbol{x}_t\right). \tag{25}$$

In this case, the MS weights do not sum to one, but it can be shown that $\text{plimit}_{T\to\infty}\sum_{t=1}^{T} w_{MS,t} = R$, where $R = pK + 1$.

Now the expected MSFE matrix from (10) takes the following vectorial form

$$\begin{aligned}\text{MSFE}(\boldsymbol{w}) &= E[\boldsymbol{e}_{\boldsymbol{w},T+1}\boldsymbol{e}_{\boldsymbol{w},T+1}'] = E\left[(\boldsymbol{x}_{T+1}'\otimes\boldsymbol{I}_K)\boldsymbol{B}\right.\\ &\left.\boldsymbol{\xi}_{T+1}\boldsymbol{\xi}_{T+1}'\boldsymbol{B}'(\boldsymbol{x}_{T+1}\otimes\boldsymbol{I}_K)\right]\\ &-(\boldsymbol{w}'\otimes\boldsymbol{I}_K)E\left[\tilde{\boldsymbol{s}}\boldsymbol{\xi}_{T+1}'\boldsymbol{B}'(\boldsymbol{x}_{T+1}\otimes\boldsymbol{I}_K)\right]\\ &-E\left[(\boldsymbol{x}_{T+1}'\otimes\boldsymbol{I}_K)\boldsymbol{B}\boldsymbol{\xi}_{T+1}\tilde{\boldsymbol{s}}'\right](\boldsymbol{w}\otimes\boldsymbol{I}_K)\\ &+\boldsymbol{\Sigma}(\boldsymbol{D}\otimes\boldsymbol{I}_K)\boldsymbol{\Sigma}'\\ &+(\boldsymbol{w}'\otimes\boldsymbol{I}_K)(\boldsymbol{\Sigma}\otimes\boldsymbol{I}_T)(\boldsymbol{D}\otimes\boldsymbol{I}_{TK})\\ &(\boldsymbol{\Sigma}'\otimes\boldsymbol{I}_T)(\boldsymbol{w}\otimes\boldsymbol{I}_K)\\ &+(\boldsymbol{w}'\otimes\boldsymbol{I}_K)E[\tilde{\boldsymbol{s}}\tilde{\boldsymbol{s}}'](\boldsymbol{w}\otimes\boldsymbol{I}_K)\end{aligned} \tag{26}$$

where

$$\tilde{\boldsymbol{s}} = \left([(\boldsymbol{x}_1'\otimes\boldsymbol{I}_K)\boldsymbol{B}\boldsymbol{\xi}_1]' \ldots [(\boldsymbol{x}_T'\otimes\boldsymbol{I}_K)\boldsymbol{B}\boldsymbol{\xi}_T]'\right)' \in \mathbb{R}^{TK}.$$

Then we obtain the following result:

Theorem 3. The unrestricted optimal weights for the $MS(M)$ $VAR(p)$ model in (19) with Markovian representation (20) are given by

$$\boldsymbol{w}_{UO}\otimes\boldsymbol{I}_K = \tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{N}} \tag{27}$$

where $\tilde{\boldsymbol{M}}$ and $\tilde{\boldsymbol{N}}$ take the following matrix expressions

$$\tilde{\boldsymbol{M}} = [\boldsymbol{\Sigma}(\boldsymbol{D}\otimes\boldsymbol{I}_K)\boldsymbol{\Sigma}']\otimes\boldsymbol{I}_T + E[\tilde{\boldsymbol{s}}\tilde{\boldsymbol{s}}'] \in \mathbb{R}^{(TK)\times(TK)} \tag{28}$$

and

$$\tilde{\boldsymbol{N}} = E\left[\tilde{\boldsymbol{s}}\boldsymbol{\xi}_{T+1}'\boldsymbol{B}'(\boldsymbol{x}_{T+1}\otimes\boldsymbol{I}_K)\right] \in \mathbb{R}^{(TK)\times K}. \tag{29}$$

Notice that $\tilde{\boldsymbol{M}}$ above is positive definite; hence, it is invertible.

Theorem 4. Let us consider the $MS(M)$ $VAR(p)$ model in (19) with a Markovian representation (20). Then the restricted optimal weights (RO) are given by

$$\boldsymbol{w}_{RO}\otimes\boldsymbol{I}_K = \tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{N}} + \frac{K-\tilde{b}}{\tilde{a}}\tilde{\boldsymbol{M}}^{-1}(\boldsymbol{i}_T\otimes\boldsymbol{I}_K) \tag{30}$$

where

$$\begin{aligned}\tilde{a} &= \text{tr}\left[(\boldsymbol{i}_T'\otimes\boldsymbol{I}_K)\tilde{\boldsymbol{M}}^{-1}(\boldsymbol{i}_T\otimes\boldsymbol{I}_K)\right]\\ \tilde{b} &= \text{tr}\left[(\boldsymbol{i}_T'\otimes\boldsymbol{I}_K)\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{N}}\right]\end{aligned}$$

and $\tilde{\boldsymbol{M}}$ and $\tilde{\boldsymbol{N}}$ are as in (28) and (29), respectively.

As before, to make the obtained formulas computationally tractable, we calculate the expectations in (26), (28), and (29) conditional $\boldsymbol{Y}_T$. Thus, one has only to apply the following relations:

$$E[\boldsymbol{\xi}_t\boldsymbol{\xi}_{T+1}'|\boldsymbol{Y}_T] = [\text{diag}(\hat{\boldsymbol{\xi}}_{t|T}) - \hat{\boldsymbol{\xi}}_{t|T}\hat{\boldsymbol{\xi}}_{t|T}']\boldsymbol{P}^{T+1-t}$$

for all $t = 1, \ldots, T$, and

$$E[\boldsymbol{\xi}_t\boldsymbol{\xi}_\tau'|\boldsymbol{Y}_T] = [\text{diag}(\hat{\boldsymbol{\xi}}_{t|T}) - \hat{\boldsymbol{\xi}}_{t|T}\hat{\boldsymbol{\xi}}_{t|T}']\boldsymbol{P}^{\tau-t}$$

for all $t, \tau = 1, \ldots, T$ with $t \le \tau$ (here we set $\boldsymbol{P}^{\tau-t} = \boldsymbol{I}_M$ for $t = \tau$). For $t > \tau$, one has to take the transpose matrix. See Appendix A for more details.

For practical inference purposes, we adopt a plug-in approach where the matrices involved in the statements of the above theorems are replaced by their ML estimates.

4 | NUMERICAL SIMULATIONS AND AN EMPIRICAL APPLICATION

4.1 | Numerical examples

We evaluate the forecast performance of the proposed weights in a series of simulated experiments. Data are generated according to MS VAR processes with $M = K = 2$, that is, a two-state bivariate AR(p) model, with $p = 0$ in the first example and $p = 1$ in the second one. The states are generated by a Markov chain whose unconditional probabilities are $\pi_1 = \pi_2 = 0.5$. The bivariate two-state MS models

$$\boldsymbol{y}_t - \boldsymbol{\Phi}_{s_t}\boldsymbol{y}_{t-1} = \boldsymbol{\nu}_{s_t} + \boldsymbol{u}_t$$

are generated with sample sizes $T = 200{,}1000$, $\boldsymbol{u}_t$ being Gaussian with zero mean and positive definite variance $\boldsymbol{\Omega}_{s_t} = \boldsymbol{\Sigma}_{s_t}\boldsymbol{\Sigma}'_{s_t}$ with $s_t \in \{1,2\}$.

In order to analyze the finite sample performances in various scenarios, we set combinations of different parameters. With regard to the intercepts, we employ the following cases:

i) $\boldsymbol{\nu}_1 = \begin{pmatrix} .2 \\ .2 \end{pmatrix} \quad \boldsymbol{\nu}_2 = \begin{pmatrix} .4 \\ .3 \end{pmatrix} \quad ||\boldsymbol{\nu}_1 - \boldsymbol{\nu}_2|| = 0.2$

ii) $\boldsymbol{\nu}_1 = \begin{pmatrix} .6 \\ .5 \end{pmatrix} \quad \boldsymbol{\nu}_2 = \begin{pmatrix} 1 \\ 1.5 \end{pmatrix} \quad ||\boldsymbol{\nu}_1 - \boldsymbol{\nu}_2|| = 1$

iii) $\boldsymbol{\nu}_1 = \begin{pmatrix} .3 \\ .2 \end{pmatrix} \quad \boldsymbol{\nu}_2 = \begin{pmatrix} 1.4 \\ 1.8 \end{pmatrix} \quad ||\boldsymbol{\nu}_1 - \boldsymbol{\nu}_2|| = 2$

where the Euclidean norm of the intercept difference increases among regimes.

With regard to the autoregressive matrices, the parameters are defined as follows:

i) $\boldsymbol{\Phi}_1 = \begin{pmatrix} .2 & .4 \\ .3 & .2 \end{pmatrix} \quad \boldsymbol{\Phi}_2 = \begin{pmatrix} .8 & .4 \\ .1 & .7 \end{pmatrix} \quad \rho(\boldsymbol{\Phi}_1)/\rho(\boldsymbol{\Phi}_2) = .5$

ii) $\boldsymbol{\Phi}_1 = \begin{pmatrix} .2 & .4 \\ .4 & .2 \end{pmatrix} \quad \boldsymbol{\Phi}_2 = \begin{pmatrix} .4 & .2 \\ .1 & .4 \end{pmatrix} \quad \rho(\boldsymbol{\Phi}_1)/\rho(\boldsymbol{\Phi}_2) = 1$

iii) $\boldsymbol{\Phi}_1 = \begin{pmatrix} .8 & .1 \\ .5 & .8 \end{pmatrix} \quad \boldsymbol{\Phi}_2 = \begin{pmatrix} .3 & .2 \\ .1 & .4 \end{pmatrix} \quad \rho(\boldsymbol{\Phi}_1)/\rho(\boldsymbol{\Phi}_2) = 2$

in which we consider equally persistent regimes in case (ii) and a much persistent regime in state 2 (case i) or in state 1 (case iii). Here $\rho(\cdot)$ denotes the spectral radius of the considered square matrix. With regard to the variance matrices, we set the following:

i) $\boldsymbol{\Sigma}_1 = \begin{pmatrix} .8 & .2 \\ .2 & .8 \end{pmatrix} \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 1 & .8 \\ .8 & 1 \end{pmatrix} \quad \rho(\boldsymbol{\Sigma}_1)/\rho(\boldsymbol{\Sigma}_2) = .5$

ii) $\boldsymbol{\Sigma}_1 = \begin{pmatrix} .8 & .2 \\ .2 & .8 \end{pmatrix} \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} .5 & .5 \\ .5 & .5 \end{pmatrix} \quad \rho(\boldsymbol{\Sigma}_1)/\rho(\boldsymbol{\Sigma}_2) = 1$

iii) $\boldsymbol{\Sigma}_1 = \begin{pmatrix} 1.9 & .9 \\ .9 & 2.3 \end{pmatrix} \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 1.3 & .2 \\ .2 & 1.3 \end{pmatrix}$

$\rho(\boldsymbol{\Sigma}_1)/\rho(\boldsymbol{\Sigma}_2) = 2$

where the volatility increases in the regimes.

The bivariate two-state processes are obtained from the above model specification by using different combinations of the true parameters. Then we determine the ratios of the MSFEs computed by using optimal unrestricted forecast weights and standard MS forecast weights, that is, $Ratio_1 = MSFE(\boldsymbol{w}_{UO})/MSFE(\boldsymbol{w}_{MS})$ or the restricted version, that is, $Ratio_2 = MSFE(\boldsymbol{w}_{RO})/MSFE(\boldsymbol{w}_{MS})$. Two facts are preliminary observed in typical realizations. First, when the modulus of the intercept difference increases, the gain using optimal weights tends to be larger (see a typical case in Figure 1). This means that, as the two regimes become more separated, the forecast is more accurate. Second, the forecast error tends to be smaller when one of the regimes is more persistent compared with the other (see a typical case in Figure 2). With such different scenarios, we run 5000 rep-

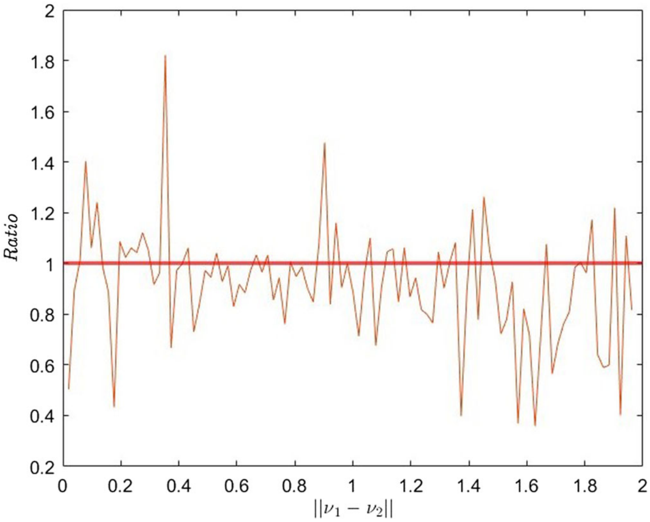

FIGURE 1 Evolution of the $Ratio_1 = MSFE(\boldsymbol{w}_{UO})/MSFE(\boldsymbol{w}_{MS})$ for increasing intercept differences in a typical realization with $T = 200$. MSFE, mean square forecast error.

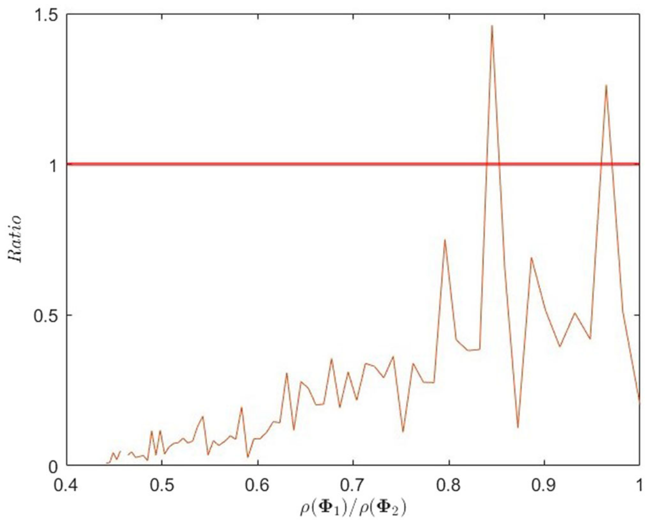

FIGURE 2 $Ratio_1 = MSFE(\boldsymbol{w}_{UO})/MSFE(\boldsymbol{w}_{MS})$ for increasing spectral ratios of the AR matrix coefficients in a typical realization with $T = 200$. MSFE, mean square forecast error.

lications. The results are reported in Table 1 for the mean-variance switching case and in Table 2 for the mean-autoregressive-variance switching case. The

TABLE 1 The table reports the ratio of the MSFE of the optimal weights to that of the Markov switching weights in the mean-variance switching case.

		$\|\|\nu_1 - \nu_2\|\|$		
	$\rho(\Sigma_1)/\rho(\Sigma_2)$	**0.2**	**1**	**2**
$T = 200$	0.5	0.9886	0.9401	0.9056
		0.9715	0.9978	0.9986
	1	1.0242	0.9691	0.9576
		0.9965	1.1646	1.0981
	2	0.9923	1.0371	1.0105
		0.9939	1.0591	1.0043
$T = 1000$	0.5	0.9873	0.8724	0.8344
		0.9599	0.9560	0.8855
	1	0.9869	0.8644	0.8595
		0.9880	0.9712	0.9375
	2	0.9927	0.9925	0.9452
		0.9946	0.9966	0.9712

Note: The first number in each cell is the unrestricted version $Ratio_1 = MSFE(\mathbf{w}_{UO})/MSFE(\mathbf{w}_{MS})$, and the second number is the restricted version $Ratio_2 = MSFE(\mathbf{w}_{RO})/MSFE(\mathbf{w}_{MS})$. The sample size is $T = 200{,}1000$, and the number of replications is 5000.
Abbreviation: MSFE, mean square forecast error.

TABLE 2 The table reports the ratio of the MSFE of the optimal weights to that of the Markov switching weights in the mean-autoregressive-variance switching case.

		$\rho(\Phi_1)/\rho(\Phi_2)$		
	$\rho(\Sigma_1)/\rho(\Sigma_2)$	**0.5**	**1**	**2**
$T = 200$	0.5	0.1735	0.3825	0.1097
		0.2008	0.4059	0.1025
	1	0.3438	0.8443	0.2564
		0.3869	0.8795	0.2219
	2	0.3960	0.9130	0.4004
		0.3980	0.9120	0.3902
$T = 1000$	0.5	0.0721	0.1562	0.0462
		0.0791	0.1619	0.0421
	1	0.1428	0.3440	0.1070
		0.1463	0.3441	0.0850
	2	0.1781	0.3801	0.1797
		0.1777	0.3823	0.1782

Note: The first number in each cell is the unrestricted version $Ratio_1 = MSFE(\mathbf{w}_{UO})/MSFE(\mathbf{w}_{MS})$, and the second number is the restricted version $Ratio_2 = MSFE(\mathbf{w}_{RO})/MSFE(\mathbf{w}_{MS})$. The sample size is $T = 200{,}1000$, and the number of replications is 5000.

general result from simulations shows that optimal weight forecasts outperform traditional forecasts in terms of accuracy. Moreover, forecasts in mean-switching case (reported in Table 1) are more precise if one of the following conditions occurs: (1) the regimes are much different in levels; (2) the variability is limited; or (3) the sample size increases. With regard to the mean-autoregressive-switching case, (reported in Table 2) forecasts are actually much more accurate. Furthermore, we observe better forecasts if one of the following cases occurs: (1) one regime is more persistent with respect to the other (for example, $\rho(\Phi_1)/\rho(\Phi_2) = .5$ or $\rho(\Phi_1)/\rho(\Phi_2) = 2$); (2) the variability is limited; or (3) the sample size increases.

4.2 | An empirical application

We analyze the relationship between the short-term interest rate and the term spread using monthly US data taken from Fred database in the period from January 1970 to October 2023. In the evaluation of the linkages between the short-term interest rate (defined as 3 month Treasury Bill rate) and the term spread (defined as 10-year government bond minus 3-month Treasury Bill rate), we take into account fluctuations of the business cycle by using a MS VAR model. The series in levels together with the first differences of the short-tem interest rate are depicted in Figure 3. Particularly, we would like firstly to evaluate whether the term spread predicts changes in the short-term interest rate when regimes are taken into account and secondly to investigate the out-of-sample forecast performance using optimal weights. We estimate a MS(2) VAR (1) model via the modified EM algorithm described in Cavicchioli (2014b) by using the obtained matrix expressions of the ML estimates of model parameters. Model selection has been performed by applying the method proposed in Cavicchioli (2014a), which is based on stable VARMA representations of a finite set of model candidates. This analysis reveals that the specification fitting well the data is in fact a MS(2) VAR(1). ML estimates of the model parameters are reported in Table 3. Standard errors in parenthesis have been computed by using the asymptotic variance-covariance matrix given in Theorem 3 from Cavicchioli (2021), which is proved to be asymptotically consistent and efficient even when the disturbances of the model are heteroskedastic and autocorrelated. Equivalently, one can use the asymptotic information matrix of the parameter estimates described in Corollary 2 from Cavicchioli (2014b). Such standard errors are robust in the sense of White (1980). This is essential for the construction of asymptotic confidence intervals and hypothesis tests.

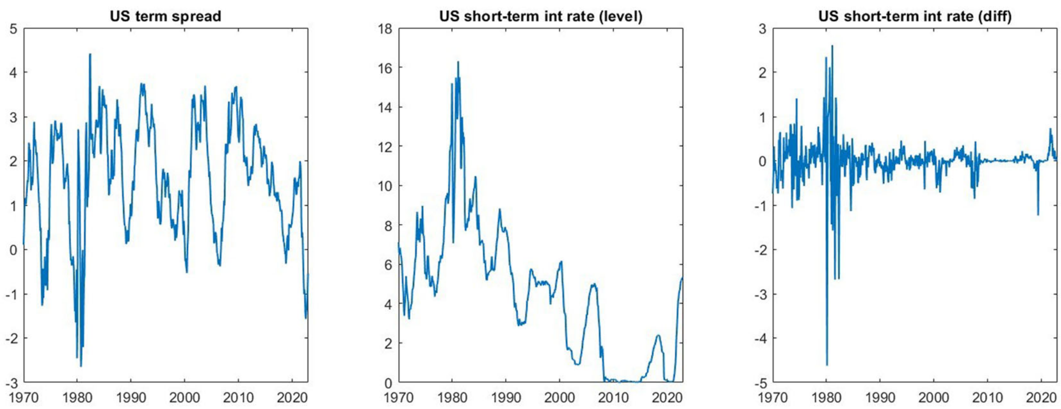

FIGURE 3 Monthly US term spread (left plot), short-term interest rate in level (central plot), and in first differences (right plot) from January 1970 to October 2023. Data are taken from FRED database.

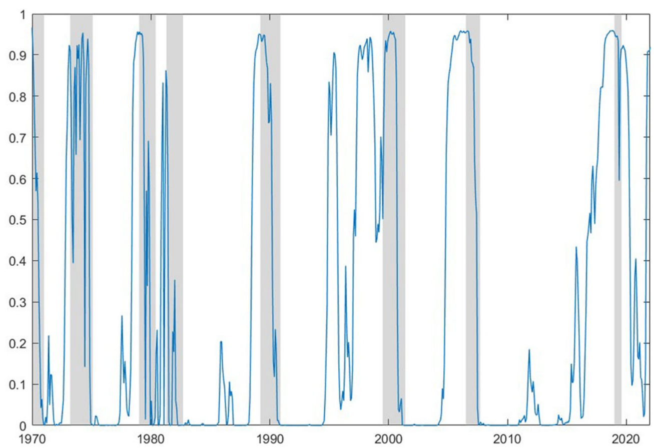

FIGURE 4 Estimated smoothed probability of regime 2 for the MS(2) VAR(1) model of monthly US term spread and short-term interest rate from January 1970 to October 2023. Shaded areas identify periods dated as recessions by the NBER. Data are taken from FRED database.

TABLE 3 ML estimates for the MS(2) VAR(1) model of monthly US term spread and short-term interest rate from January 1970 to October 2023.

	Regime 1 (E)		***Regime 2 (R)***	
$\boldsymbol{\nu}$	.062		.169	
	(001.)		(.046)	
	−.039		−.135	
	(.001)		(.055)	
$\boldsymbol{\Phi}$	.919	−.327	.834	−.106
	(.099)	(.021)	(.035)	(.015)
	.063	.446	.113	.068
	(.012)	(.024)	(.043)	(.018)
$\boldsymbol{\Sigma}$	.070	−.054	.129	−.115
	(.006)	(.005)	(.073)	(.059)
	−.054	.097	−.115	.195
	(.005)	(.004)	(.059)	(.048)
	$p_{11} = .69$		$p_{22} = .31$	

Note: Data are taken from FRED database. Standard errors in parenthesis.
Abbreviation: ML, maximum likelihood.

Alternatively, one can employ the asymptotic heteroskedasticity and autocorrelation consistent covariance matrices derived in Theorem 1 from White (1980) and Theorem 2 from Newey and West (1987), respectively. The estimated smoothed probabilities of regime 2 are depicted in Figure 4, where shaded areas identify periods dated as recessions by the National Bureau of Economic Research (NBER). We identify a second regime where there exist a higher spread and a more negative interest rate compared to regime 1 together with a higher volatility. On the contrary, regime 1 exhibits higher persistence in the dynamics of the time series and a higher probability of remaining in such a regime. For this reason, we recognize regime 1 being normal or expansionary times and regime 2 being turbulent or recessionary periods. This is also confirmed by the estimated smoothed probabilities that coincide with the major US crisis. The out-of sample forecasts are reported in Table 4 along with the Diebold and Mariano (1995) test for equal accuracy in prediction. The test is employed by accounting for heteroskedasticity and autocorrelation usually affecting multiperiod forecast errors. We conclude that optimal forecasts improve the MS ones in both full sample and the selected subsamples. Moreover, the performance is superior in the unrestricted version, and the test indicates significantly higher accuracy in forecasts.

TABLE 4 Out-of-sample forecast performances.

	$MSFE(w_{UO})/MSFE(w_{MS})$	$MSFE(w_{RO})/MSFE(w_{MS})$
Full sample		
January 2003 to October 2023	0.163**	0.589**
Subperiods		
January 2003 to January 2008	0.266**	0.624*
February 2008 to January 2013	0.198**	0.511**
February 2013 to January 2018	0.345*	0.643*
February 2018 to October 2023	0.175**	0.561**

Abbreviation: MSFE, mean square forecast error.
*Denotes significance of the test at 5%.
**Denotes significance of the test at 1%.

5 | CONCLUDING REMARKS

In this paper, we derive optimal forecasts for multivariate MS autoregressive models obtained by minimizing the trace of the expected MSFE matrix. Then we analyze the effect of uncertainty around states comparing forecasts based on MS and unrestricted (resp. restricted) optimal weights conditional on available observables. The key to the solution is to use suitable state space representations of the considered MS models. Starting from such representations, we derive explicit neat matrix expressions in closed form for the forecasting optimal weights in terms of the matrices involved in the state space representation of the specified model. So we use quantities already available by standard estimation of the model parameters. The obtained formulas are easily tractable and programmable. The results show that optimal forecasts can differ substantially from standard MS forecasts. This also confirms in the multivariate setting the issues derived in Boot and Pick (2018) for univariate MS AR(0) models with exogeneous regressors. Simulations and an empirical application illustrate the feasibility of the proposed approach and show that forecasts with optimal weights outperform the traditional MS weights. This approach is quite easy to work with and can be used to forecast for more sophisticated MS VAR (co-integrated) processes, as listed in Krolzig ((1997), §13, 306–310). It is also worth mentioning that a possible line of further research could be to study optimal forecasts for MS Bilinear processes, introduced and considered in Bibi and Ghezal (2015).

As stated in Section 1, MS models are usually outperformed by linear models. In this sense, the referee suggests that the Model Confidence Set (MCS), introduced and studied in Hansen et al. (2011), might be useful in evaluating the out-of-sample performance obtained using optimal forecast weights relative to both traditional MS weights and linear models. The MCS procedure yields a model confidence set that is a collection of models built to contain the best model(s) with a given level of confidence. A bootstrap method to implement the MCS for comparisons of forecasting models evaluated out-of-sample has been proposed in Section 3 of the cited paper. We leave to future research the use of this interesting issue for investigating the performance of our approach.

ACKNOWLEDGMENTS

This work was financially supported by a FAR 2022 research grant of the University of Modena and Reggio E., Italy. We would like to thank the Editor of the journal, Professor Derek Bunn, and the anonymous referee for his/her constructive comments and suggestions which were most valuable for improvement of the final version of the paper. Open access publishing facilitated by Universita degli Studi di Modena e Reggio Emilia, as part of the Wiley - CRUI-CARE agreement.

CONFLICT OF INTEREST STATEMENT

The authors have no relevant financial or nonfinancial interests to disclose.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Maddalena Cavicchioli https://orcid.org/0000-0001-7125-8111

REFERENCES

Alemohammad, N., Rezakhah, S., & Alizadeh, S. H. (2020). Markov switching asymmetric GARCH model: Stability and forecasting. *Statistical Papers*, *61*(3), 1309–1333.

Alvarez, R., Camacho, M., & Ruiz, M. (2017). Inference on filtered and smoothed probabilities in Markov-switching autoregressive models. *Journal of Business & Economic Statistics*, *37*(3), 484–495.

Ang, A., & Bekaert, G. (2002). Regime switches in interest rates. *Journal of Business & Economic Statistics*, *20*(2), 163–182.

Ardia, D., Bluteau, K., Boudt, K., & Catania, L. (2018). Forecasting risk with Markov-switching GARCH models: A large-scale performance study. *International Journal of Forecasting*, *34*, 733–747.

Bessec, M., & Bouabdallah, O. (2005). What causes the forecasting failure of Markov-switching models? A Monte Carlo study. *Studies in Nonlinear Dynamics & Econometrics*, *9*(2).

Bibi, A., & Ghezal, A. (2015). On the Markov-switching bilinear processes: Stationarity, higher-order moments and β-mixing. *Stochastics. An International Journal of Probability and Stochastic Processes*, *87*(6), 919–945.

Boot, T., & Nibbering, D. (2019). Forecasting using random subspace methods. *Journal of Econometrics*, *209*(2), 391–406.

Boot, T., & Pick, A. (2018). Optimal forecasts from Markov switching models. *Journal of Business & Economic Statistics*, *36*(4), 628–642.

Boot, T., & Pick, A. (2020). Does modeling a structural break improve forecast accuracy? *Journal of Econometrics*, *215*(1), 35–59.

Cavicchioli, M. (2013). Spectral density of Markov-switching VARMA models. *Economics Letters*, *121*, 218–220.

Cavicchioli, M. (2014a). Determining the number of regimes in Markov–switching *VAR* and *VMA* models. *Journal of Time Series Analysis*, *35*(2), 173–186.

Cavicchioli, M. (2014b). Analysis of the likelihood function for Markov switching VAR(CH) models. *Journal of Time Series Analysis*, *35*(6), 624–639.

Cavicchioli, M. (2016). Statistical analysis of mixture vector autoregressive models. *Scandinavian Journal of Statistics*, *43*(4), 1192–1213.

Cavicchioli, M. (2017a). Higher order moments of Markov switching VARMA models. *Econometric Theory*, *33*(6), 1502–1515.

Cavicchioli, M. (2017b). Asymptotic Fisher information matrix of Markov switching VARMA models. *Journal of Multivariate Analysis*, *157*, 124–135.

Cavicchioli, M. (2021). OLS estimation Of Markov switching VAR models: Asymptotics and application to energy use. *AStA Advances in Statistical Analysis*, *105*, 431–449.

Cavicchioli, M. (2022). Goodness-of-fit tests for Markov switching VAR models using spectral analysis. *Journal of Statistical Planning and Inference*, *219*, 189–203.

Chen, C. W. S., So, M. K. P., & Lin, E. M. H. (2009). Volatility forecasting with double Markov-switching GARCH models. *Journal of Forecasting*, *28*(8), 681–697.

Clements, M. P., & Krolzig, H. M. (1998). A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP. *Econometrics Journal*, *1*, 47–75.

Clements, M. P., Franses, P. H., & Swanson, N. R. (2004). Forecasting economic and financial time-series with non-linear models. *International Journal of Forecasting*, *20*(2), 169–183.

Dacco, R., & Satchell, S. (1999). Why do regime-switching models forecast so badly. *Journal of Forecasting*, *18*, 1–16.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263.

Douc, R., Moulines, E., & Ryden, T. (2004). Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *Annals of Statistics*, *32*, 2254–2304.

Elliott, G., & Timmermann, A. (2005). Optimal forecast combination under regime switching. *International Economic Review*, *46*(4), 1081–1102.

Foroni, C., Guérin, P., & Marcellino, M. (2015). Markov-switching mixed-frequency VAR models. *International Journal of Forecasting*, *31*, 692–711.

Francq, C., & Zakoïan, J. M. (2001). Stationarity of multivariate Markov switching ARMA models. *Journal of Econometrics*, *102*, 339–364.

Guidolin, M., & Timmermann, A. (2009). Forecasts of US short-term interest rates: A flexible forecast combination approach. *Journal of Econometrics*, *150*(2), 297–311.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, *57*, 357–384.

Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics*, *45*, 39–70.

Hamilton, J. D. (1994). *Time series analysis*: Princeton University Press.

Hamilton, J. D. (2016). Chapter 3-macroeconomic regimes and regime shifts. *Handbook of macroeconomics* (Vol. 2). Holland: Elsevier.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.

Hou, C. (2017). Infinite hidden Markov switching VARs with application to macroeconomic forecast. *International Journal of Forecasting*, *33*(4), 1025–1043.

Knuppel, M. (2009). Testing business cycle asymmetries based on autoregressions with a Markov-switching intercept. *Journal of Business & Economic Statistics*, *27*(4), 544–552.

Kontolemis, Z. G. (2001). Analysis of the US business cycle with a vector-Markov-switching model. *Journal of Forecasting*, *20*, 47–61.

Krolzig, H. M. (1997). Markov-switching vector autoregressions: Modelling, statistical inference, and application to business cycle analysis. *Lecture Notes in Economics and Mathematical Systems*. Berlin-Heidelberg-New York: Springer-Verlag.

Krolzig, H. M. (2000). Predicting Markov-switching vector autoregressive processes. (*Working Paper Department of Economics and Nuffield College*). Oxford, UK.

Kundu, S., & Paul, A. (2022). Effect of economic policy uncertainty on stock market return and volatility under heterogeneous market characteristics. *International Review of Economic and Finance*, *80*, 597–612.

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708.

Nikolsko-Rzhevskyy, A., & Prodan, R. (2012). Markov switching and exchange rate predictability. *International Journal of Forecasting*, *28*(2), 353–365.

Nyberg, H. (2018). Forecasting US interest rates and business cycle with a nonlinear regime switching VAR model. *Journal of Forecasting*, *37*(1), 1–15.

Otranto, E. (2005). The multi-chain Markov switching model. *Journal of Forecasting*, *24*(7), 523–537.

Parikakis, G. S., & Merika, A. (2009). Evaluating volatility dynamics and the forecasting ability of Markov switching models. *Journal of Forecasting*, *28*(8), 736–744.

Pesaran, H. M., Pettenuzzo, D., & Timmermann, A. (2006). Forecasting time series subject to multiple structural breaks. *Review of Economic Studies*, *73*, 1057–1084.

Prüser, J. (2021). Forecasting US inflation using Markov dimension switching. *Journal of Forecasting*, *40*(3), 481–499.

Stelzer, R. (2009). On Markov-switching ARMA processes–stationarity, existence of moments and geometric ergodicity. *Econometric Theory*, *25*(1), 43–62.

White, H. (1980). A heteroskedasticity consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*(4), 817–838.

AUTHOR BIOGRAPHY

Maddalena Cavicchioli is an Associate Professor of Economic Statistics at the Department of Economics Marco Biagi, University of Modena and Reggio Emilia.

Her main research interests are in the areas of (i) time series analysis with particular focus on nonlinear models and their estimation algorithm, (ii) dimension reduction methods and statistical learning techniques for large datasets, and (iii) statistical modelling of economic and financial time series and managerial data. She is the author of high profile papers in well-known international journals such as Journal of Business & Economic Statistics, Econometric Theory, Journal of Financial Econometrics, Electronic Journal of Statistics, Journal of Statistical Planning and Inference, Scandinavian Journal of Statistics, Journal of Multivariate Analysis, AStA Advances in Statistical Analysis, Computational Statistics, Statistical Methods & Applications, Statistical Papers and Corporate Social Responsibility and Environmental Management. She presents her works in numerous national and international conferences and involved in various national and international research groups.

How to cite this article: Cavicchioli, M. (2024). Forecasting Markov switching vector autoregressions: Evidence from simulation and application. *Journal of Forecasting*, 1–17. https://doi.org/10.1002/for.3180

APPENDIX A

I) *Derivation of* (12) *and further computations*. We have

$$\begin{aligned}\boldsymbol{e}_{\boldsymbol{w},T+1} &= \boldsymbol{y}_{T+1} - \hat{\boldsymbol{y}}_{\boldsymbol{w},T+1} = \boldsymbol{\Lambda}\boldsymbol{\xi}_{T+1} \\ &+ \boldsymbol{\Sigma}\left(\boldsymbol{\xi}_{T+1} \otimes \boldsymbol{I}_K\right)\boldsymbol{u}_{T+1} - \sum_{t=1}^{T} w_t \boldsymbol{y}_t \\ &= \boldsymbol{\Lambda}\boldsymbol{\xi}_{T+1} + \boldsymbol{\Sigma}\left(\boldsymbol{\xi}_{T+1} \otimes \boldsymbol{I}_K\right)\boldsymbol{u}_{T+1} \\ &- \sum_{t=1}^{T} w_t [\boldsymbol{\Lambda}\boldsymbol{\xi}_t + \boldsymbol{\Sigma}(\boldsymbol{\xi}_t \otimes \boldsymbol{I}_K)\boldsymbol{u}_t] \\ &= \boldsymbol{\Lambda}\left[\boldsymbol{\xi}_{T+1} - \sum_{t=1}^{T} w_t \boldsymbol{\xi}_t\right] + \boldsymbol{\Sigma}\left(\boldsymbol{\xi}_{T+1} \otimes \boldsymbol{u}_{T+1}\right) \\ &- \boldsymbol{\Sigma}\sum_{t=1}^{T} w_t (\boldsymbol{\xi}_t \otimes \boldsymbol{u}_t) \\ &\boldsymbol{\Lambda}\left[\boldsymbol{\xi}_{T+1} - \left(\boldsymbol{w}' \otimes \boldsymbol{I}_M\right)\boldsymbol{s}\right] + \boldsymbol{\Sigma}\left(\boldsymbol{\xi}_{T+1} \otimes \boldsymbol{u}_{T+1}\right) \\ &- \boldsymbol{\Sigma}\sum_{t=1}^{T} w_t (\boldsymbol{\xi}_t \otimes \boldsymbol{u}_t)\end{aligned}$$

as $(\boldsymbol{w}' \otimes \boldsymbol{I}_M)\boldsymbol{s} = \sum_{t=1}^{T} w_t \boldsymbol{\xi}_t$. Then the MSFE matrix is given by

$$\begin{aligned}&E\left[\boldsymbol{e}_{\boldsymbol{w},T+1}\boldsymbol{e}'_{\boldsymbol{w},T+1}\right] = \boldsymbol{\Lambda}E\left[\left(\boldsymbol{\xi}_{T+1} - (\boldsymbol{w}' \otimes \boldsymbol{I}_M)\boldsymbol{s}\right)\right. \\ &\left.\left(\boldsymbol{\xi}'_{T+1} - \boldsymbol{s}'(\boldsymbol{w} \otimes \boldsymbol{I}_M)\right)\right]\boldsymbol{\Lambda}' \\ &+ \boldsymbol{\Sigma}E\left[(\boldsymbol{\xi}_{T+1}\boldsymbol{\xi}'_{T+1}) \otimes (\boldsymbol{u}_{T+1}\boldsymbol{u}'_{T+1})\right]\boldsymbol{\Sigma}' \\ &+ \boldsymbol{\Sigma}\sum_{t=1}^{T} w_t^2 E[\boldsymbol{\xi}_t\boldsymbol{\xi}'_t] \otimes E[\boldsymbol{u}_t\boldsymbol{u}'_t]\boldsymbol{\Sigma}' \\ &= \boldsymbol{\Lambda}E[\boldsymbol{\xi}_{T+1}\boldsymbol{\xi}'_{T+1}]\boldsymbol{\Lambda}' - \boldsymbol{\Lambda}E[\boldsymbol{\xi}_{T+1}\boldsymbol{s}'](\boldsymbol{\Lambda}' \otimes \boldsymbol{I}_T)(\boldsymbol{w} \otimes \boldsymbol{I}_K) \\ &- (\boldsymbol{w}' \otimes \boldsymbol{I}_K)(\boldsymbol{\Lambda} \otimes \boldsymbol{I}_T)E[\boldsymbol{s}\boldsymbol{\xi}'_{T+1}]\boldsymbol{\Lambda}' \\ &+ (\boldsymbol{w}' \otimes \boldsymbol{I}_K)(\boldsymbol{\Lambda} \otimes \boldsymbol{I}_T)E[\boldsymbol{s}\boldsymbol{s}'](\boldsymbol{\Lambda}' \otimes \boldsymbol{I}_T)(\boldsymbol{w} \otimes \boldsymbol{I}_K) \\ &+ \boldsymbol{\Sigma}\{E[\boldsymbol{\xi}_{T+1}\boldsymbol{\xi}'_{T+1}] \otimes E[\boldsymbol{u}_{T+1}\boldsymbol{u}'_{T+1}]\}\boldsymbol{\Sigma}' \\ &+ (\boldsymbol{w}' \otimes \boldsymbol{I}_K)(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_T) \\ &\left\{E\left[\boldsymbol{\xi}_t\boldsymbol{\xi}'_t\right] \otimes E[\boldsymbol{u}_t\boldsymbol{u}'_t] \otimes \boldsymbol{I}_T\right\}(\boldsymbol{\Sigma}' \otimes \boldsymbol{I}_T)(\boldsymbol{w} \otimes \boldsymbol{I}_K)\end{aligned}$$

which implies (12). Here we have used $E[\boldsymbol{\xi}_t\boldsymbol{u}'_\tau] = \boldsymbol{0}$, for all t and τ, $E[\boldsymbol{u}_t\boldsymbol{u}'_\tau] = \boldsymbol{0}$ for all $t \neq \tau$, $E\left[\boldsymbol{\xi}_t\boldsymbol{\xi}'_t\right] = \boldsymbol{D}$ and $E[\boldsymbol{u}_t\boldsymbol{u}'_t] = \boldsymbol{I}_K$, for all t.

Furthermore, we get the following block matrix

$$\begin{aligned}E[\boldsymbol{\xi}_{T+1}\boldsymbol{s}'] &= \left[E[\boldsymbol{\xi}_{T+1}\boldsymbol{\xi}'_1] \quad E[\boldsymbol{\xi}_{T+1}\boldsymbol{\xi}'_2] \ldots E[\boldsymbol{\xi}_{T+1}\boldsymbol{\xi}'_T]\right] \\ &= \left[(\boldsymbol{P}')^T\boldsymbol{D} \quad (\boldsymbol{P}')^{T-1}\boldsymbol{D} \ldots \boldsymbol{P}'\boldsymbol{D}\right] \in \mathbb{R}^{M \times (TM)}\end{aligned}$$

as $E[\boldsymbol{\xi}_t\boldsymbol{\xi}'_{t+h}] = \boldsymbol{D}\boldsymbol{P}^h$, for all $h > 0$, hence $E[\boldsymbol{\xi}_{t+h}\boldsymbol{\xi}'_t] = (\boldsymbol{P}')^h\boldsymbol{D}$. Then the transpose block matrix is

$$E[\boldsymbol{s}\boldsymbol{\xi}'_{T+1}] = \begin{pmatrix} \boldsymbol{D}\boldsymbol{P}^T \\ \boldsymbol{D}\boldsymbol{P}^{T-1} \\ \vdots \\ \boldsymbol{D}\boldsymbol{P} \end{pmatrix} \in \mathbb{R}^{(TM)\times M}.$$

Finally, we obtain

$$\begin{aligned} E[\boldsymbol{s}\boldsymbol{s}'] &= \begin{pmatrix} E[\boldsymbol{\xi}_1\boldsymbol{\xi}'_1] & E[\boldsymbol{\xi}_1\boldsymbol{\xi}'_2] & \cdots & E[\boldsymbol{\xi}_1\boldsymbol{\xi}'_T] \\ E[\boldsymbol{\xi}_2\boldsymbol{\xi}'_1] & E[\boldsymbol{\xi}_2\boldsymbol{\xi}'_2] & \cdots & E[\boldsymbol{\xi}_2\boldsymbol{\xi}'_T] \\ \vdots & \vdots & & \vdots \\ E[\boldsymbol{\xi}_T\boldsymbol{\xi}'_1] & E[\boldsymbol{\xi}_T\boldsymbol{\xi}'_2] & \cdots & E[\boldsymbol{\xi}_T\boldsymbol{\xi}'_T] \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{D} & \boldsymbol{D}\boldsymbol{P} & \cdots & \boldsymbol{D}\boldsymbol{P}^{T-1} \\ \boldsymbol{P}'\boldsymbol{D} & \boldsymbol{D} & \cdots & \boldsymbol{D}\boldsymbol{P}^{T-2} \\ \vdots & \vdots & & \vdots \\ (\boldsymbol{P}')^{T-1}\boldsymbol{D} & (\boldsymbol{P}')^{T-2}\boldsymbol{D} & \cdots & \boldsymbol{D} \end{pmatrix}. \end{aligned}$$

Substituting the above matrix formulas into (12) yields a neat and easily programmable expression for the MSFE.

II) *Derivation of* (13)–(16). The first derivatives of the trace of MSFE($\boldsymbol{w}$) with respect to $\boldsymbol{w}\otimes\boldsymbol{I}_K$ are given by the $(TK)\times(TK)$ matrix

$$\begin{aligned} &\frac{\partial\,\mathrm{tr}\mathrm{MSFE}(\boldsymbol{w})}{\partial\boldsymbol{w}\otimes\boldsymbol{I}_K} = -2(\boldsymbol{\Lambda}\otimes\boldsymbol{I}_T)E\left[\boldsymbol{s}\boldsymbol{\xi}'_{T+1}\right]\boldsymbol{\Lambda}' \\ &+2[(\boldsymbol{\Lambda}\otimes\boldsymbol{I}_T)E[\boldsymbol{s}\boldsymbol{s}'](\boldsymbol{\Lambda}'\otimes\boldsymbol{I}_T)+(\boldsymbol{\Sigma}\otimes\boldsymbol{I}_T) \\ &(\boldsymbol{D}\otimes\boldsymbol{I}_{TK})(\boldsymbol{\Sigma}'\otimes\boldsymbol{I}_T)](\boldsymbol{w}\otimes\boldsymbol{I}_K). \end{aligned}$$

Equating to the null matrix yields

$$\boldsymbol{M}(\boldsymbol{w}_{UO}\otimes\boldsymbol{I}_K) = \boldsymbol{N}$$

where $\boldsymbol{M}$ and $\boldsymbol{N}$ are as in (14) and (15). This proves (13), and implies (16) as $\boldsymbol{M}$ is positive definite, hence invertible.

For computational purposes, we use the following neat expressions for the matrices $\boldsymbol{M} = \boldsymbol{M}_1 + \boldsymbol{M}_2$ and $\boldsymbol{N}$:

$$\begin{aligned} \boldsymbol{M}_1 &= \mathrm{diag}[\boldsymbol{\Sigma}(\boldsymbol{D}\otimes\boldsymbol{I}_K)\boldsymbol{\Sigma}' \cdots \boldsymbol{\Sigma}(\boldsymbol{D}\otimes\boldsymbol{I}_K)\boldsymbol{\Sigma}'] \in \mathbb{R}^{(TK)\times(TK)} \\ \boldsymbol{M}_2 &= (\boldsymbol{\Lambda}\otimes\boldsymbol{I}_T)E[\boldsymbol{s}\boldsymbol{s}'](\boldsymbol{\Lambda}'\otimes\boldsymbol{I}_T) \\ &= \begin{pmatrix} \boldsymbol{\Lambda}\boldsymbol{D}\boldsymbol{\Lambda}' & \boldsymbol{\Lambda}\boldsymbol{D}\boldsymbol{P}\boldsymbol{\Lambda}' & \cdots & \boldsymbol{\Lambda}\boldsymbol{D}\boldsymbol{P}^{T-1}\boldsymbol{\Lambda}' \\ \boldsymbol{\Lambda}\boldsymbol{P}'\boldsymbol{D}\boldsymbol{\Lambda}' & \boldsymbol{\Lambda}\boldsymbol{D}\boldsymbol{\Lambda}' & \cdots & \boldsymbol{\Lambda}\boldsymbol{D}\boldsymbol{P}^{T-2}\boldsymbol{\Lambda}' \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\Lambda}(\boldsymbol{P}')^{T-1}\boldsymbol{D}\boldsymbol{\Lambda}' & \boldsymbol{\Lambda}(\boldsymbol{P}')^{T-2}\boldsymbol{D}\boldsymbol{\Lambda}' & \cdots & \boldsymbol{\Lambda}\boldsymbol{D}\boldsymbol{\Lambda}' \end{pmatrix} \in \mathbb{R}^{(TK)\times(TK)} \end{aligned}$$

and

$$\boldsymbol{N} = (\boldsymbol{\Lambda} \otimes \boldsymbol{I}_T) E\left[\boldsymbol{s}\boldsymbol{\xi}'_{T+1}\right]\boldsymbol{\Lambda}' = \begin{pmatrix} \boldsymbol{\Lambda}\boldsymbol{D}\boldsymbol{P}^T\boldsymbol{\Lambda}' \\ \boldsymbol{\Lambda}\boldsymbol{D}\boldsymbol{P}^{T-1}\boldsymbol{\Lambda}' \\ \vdots \\ \boldsymbol{\Lambda}\boldsymbol{D}\boldsymbol{P}\boldsymbol{\Lambda}' \end{pmatrix} \in \mathbb{R}^{(TK)\times K}$$

III) *Derivation of* (18). To minimize the trace of $E\left[\boldsymbol{e}_{\boldsymbol{w},T+1}\boldsymbol{e}'_{\boldsymbol{w},T+1}\right]$ subject to $\sum_{t=1}^{T} w_t = 1$, that is, $\boldsymbol{i}'_T\boldsymbol{w} = 1$, we first consider the matrix Lagrangian

$$\mathcal{L}(\boldsymbol{w}) = E\left[\boldsymbol{e}_{\boldsymbol{w},T+1}\boldsymbol{e}'_{\boldsymbol{w},T+1}\right] + \lambda(1 - \boldsymbol{w}'\boldsymbol{i}_T)\boldsymbol{I}_K \in \mathbb{R}^{K\times K}$$

whose first derivatives with respect to $\boldsymbol{w}$ is the $(TK)\times K$ matrix

$$\frac{\partial\mathcal{L}(\boldsymbol{w})}{\partial\boldsymbol{w}} = \boldsymbol{M}(\boldsymbol{w}\otimes\boldsymbol{I}_K) - \boldsymbol{N} - \lambda(\boldsymbol{i}_T\otimes\boldsymbol{I}_K)$$

Equating to the null matrix gives

$$\lambda(\boldsymbol{i}_T\otimes\boldsymbol{I}_K) = \boldsymbol{M}(\boldsymbol{w}\otimes\boldsymbol{I}_K) - \boldsymbol{N} \tag{A1}$$

or, equivalently,

$$\lambda\boldsymbol{M}^{-1}(\boldsymbol{i}_T\otimes\boldsymbol{I}_K) = (\boldsymbol{w}\otimes\boldsymbol{I}_K) - \boldsymbol{M}^{-1}\boldsymbol{N}$$

hence,

$$\begin{aligned} \lambda\left(\boldsymbol{i}'_T\otimes\boldsymbol{I}_K\right)\boldsymbol{M}^{-1}(\boldsymbol{i}_T\otimes\boldsymbol{I}_K) &= \left(\boldsymbol{i}'_T\otimes\boldsymbol{I}_K\right)(\boldsymbol{w}\otimes\boldsymbol{I}_K) - \left(\boldsymbol{i}'_T\otimes\boldsymbol{I}_K\right)\boldsymbol{M}^{-1}\boldsymbol{N} \\ &= (\boldsymbol{i}'_T\boldsymbol{w})\otimes\boldsymbol{I}_K - \left(\boldsymbol{i}'_T\otimes\boldsymbol{I}_K\right)\boldsymbol{M}^{-1}\boldsymbol{N} \\ &= \boldsymbol{I}_K - \left(\boldsymbol{i}'_T\otimes\boldsymbol{I}_K\right)\boldsymbol{M}^{-1}\boldsymbol{N}. \end{aligned}$$

Taking the trace operator gives $\lambda a = K - b$; hence, $\lambda = (K-b)/a \in \mathbb{R}$, where a and b are as in §2 (after formula 18). Substituting λ into (A1) yields

$$\boldsymbol{M}(\boldsymbol{w}_{RO}\otimes\boldsymbol{I}_K) - \boldsymbol{N} - \frac{K-b}{a}(\boldsymbol{i}_T\otimes\boldsymbol{I}_K) = \boldsymbol{0}$$

which implies (18).

IV) *Derivation of* (25). Substituting (21) into (24) yields

$$
\begin{aligned}
\hat{\boldsymbol{y}}_{T+1|T} &= \sum_{i=1}^{M} \hat{\xi}_{i,T+1|T} \left(\boldsymbol{x}'_{T+1} \otimes \boldsymbol{I}_K\right) \left(\boldsymbol{X}_{iT}^{-1} \otimes \boldsymbol{I}_K\right) \left[\sum_{t=1}^{T} (\boldsymbol{x}_t \otimes \boldsymbol{I}_K) \boldsymbol{y}_t \hat{\xi}_{it|T}\right] \\
&= \sum_{i=1}^{M} \hat{\xi}_{i,T+1|T} \left[\left(\boldsymbol{x}'_{T+1} \boldsymbol{X}_{iT}^{-1}\right) \otimes \boldsymbol{I}_K\right] \left[\sum_{t=1}^{T} (\boldsymbol{x}_t \otimes \boldsymbol{I}_K) \boldsymbol{y}_t \hat{\xi}_{it|T}\right] \\
&= \sum_{t=1}^{T} \sum_{i=1}^{M} \hat{\xi}_{i,T+1|T} \hat{\xi}_{it|T} \left[\left(\boldsymbol{x}'_{T+1} \boldsymbol{X}_{iT}^{-1} \boldsymbol{x}_t\right) \otimes \boldsymbol{I}_K\right] \boldsymbol{y}_t \\
&= \sum_{t=1}^{T} \left[\sum_{i=1}^{M} \hat{\xi}_{i,T+1|T} \hat{\xi}_{it|T} \left(\boldsymbol{x}'_{T+1} \boldsymbol{X}_{iT}^{-1} \boldsymbol{x}_t\right)\right] \boldsymbol{y}_t \\
&= \sum_{t=1}^{T} w_{MS,t} \boldsymbol{y}_t
\end{aligned}
$$

which implies (25).

V) *On the effective computations of* (27)–(30). It suffices to derive the (t, τ) block matrix of $E[\tilde{\boldsymbol{s}}\tilde{\boldsymbol{s}}'] \in \mathbb{R}^{(TK)\times(TK)}$, for all $t, \tau = 1, \ldots, T$, $t \leq \tau$ (by symmetry), and the $(t, 1)$ block matrix of $E\left[\tilde{\boldsymbol{s}}\boldsymbol{\xi}'_{T+1}\boldsymbol{B}'(\boldsymbol{x}_{T+1} \otimes \boldsymbol{I}_K)\right] \in \mathbb{R}^{(TK)\times K}$, for all $t = 1, \ldots, T$. By the law of iterated expectations, we can use the following approximations for T sufficiently large:

$$
\begin{aligned}
& E[\tilde{\boldsymbol{s}}\tilde{\boldsymbol{s}}']_{t\tau} = E[E[\tilde{\boldsymbol{s}}\tilde{\boldsymbol{s}}' | \boldsymbol{Y}_T]]_{t\tau} \\
& = E\left[\left(\boldsymbol{x}'_t \otimes \boldsymbol{I}_K\right) \boldsymbol{B} E\left[\boldsymbol{\xi}_t \boldsymbol{\xi}'_\tau | \boldsymbol{Y}_T\right] \boldsymbol{B}' (\boldsymbol{x}_\tau \otimes \boldsymbol{I}_K)\right]_{t\tau} \\
& \sim \frac{1}{T^2} \sum_{t=1}^{T} \sum_{\tau=t}^{T} \left(\boldsymbol{x}'_t \otimes \boldsymbol{I}_K\right) \hat{\boldsymbol{B}} \\
& \left[\operatorname{diag}(\hat{\boldsymbol{\xi}}_{t|T}) - \hat{\boldsymbol{\xi}}_{t|T} \hat{\boldsymbol{\xi}}'_{t|T}\right] \boldsymbol{P}^{\tau - t} \hat{\boldsymbol{B}}' (\boldsymbol{x}_\tau \otimes \boldsymbol{I}_K) \in \mathbb{R}^{K \times K}
\end{aligned}
$$

$$
\begin{aligned}
& E[\tilde{\boldsymbol{s}}\boldsymbol{\xi}'_{T+1} \boldsymbol{B}' (\boldsymbol{x}_{T+1} \otimes \boldsymbol{I}_K)]_{t1} \\
& = E\left[\left(\boldsymbol{x}'_t \otimes \boldsymbol{I}_K\right) \boldsymbol{B} E[\boldsymbol{\xi}_t \boldsymbol{\xi}'_{T+1} | \boldsymbol{Y}_T] \boldsymbol{B}' (\boldsymbol{x}_{T+1} \otimes \boldsymbol{I}_K)\right]_{t1} \\
& \sim \frac{1}{T} \sum_{t=1}^{T} \left(\boldsymbol{x}'_t \otimes \boldsymbol{I}_K\right) \hat{\boldsymbol{B}} \left[\operatorname{diag}(\hat{\boldsymbol{\xi}}_{t|T}) - \hat{\boldsymbol{\xi}}_{t|T} \hat{\boldsymbol{\xi}}'_{t|T}\right] \\
& \boldsymbol{P}^{T+1-t} \hat{\boldsymbol{B}}' (\boldsymbol{x}_{T+1} \otimes \boldsymbol{I}_K) \in \mathbb{R}^{K \times K}.
\end{aligned}
$$