Statistics and Computing (2026) 36:154
https://doi.org/10.1007/s11222-026-10897-7

ORIGINAL PAPER

Multivariate Markov switching BEKK models: filtering, estimation and data analysis

Maddalena Cavicchioli[1] · Jie Cheng[2]

Received: 5 February 2025 / Accepted: 29 April 2026

Abstract

This paper deals with an extension of the standard multivariate BEKK model, as detailed in Engle and Kroner (1995), by allowing both the unconditional correlation and the parameters to be driven by an unobservable Markov chain. We propose two estimation algorithms by using extended Kalman filters, derived from suitable state space representations of the considered model. Numerical examples make evident the effectiveness of the proposed nonlinear estimations. Moreover, real-data applications on some financial returns show empirical evidence that the high volatility persistence and correlation changes of such returns can be well explained by estimating multivariate Markov switching BEKK parameters via the two efficient proposed algorithms. Finally, such results are compared to those obtained using Markov switching CCC and DCC models from Billio and Caporin (2005) to analyze financial contagion in the stock market and value-at-risk forecasts.

Keywords Markov switching models · Markov switching BEKK models · State space representations · Extended Kalman filter · Parameter estimation · Volatility · Financial returns

JEL Classification C32 · G17 · G32

1 Introduction

Financial data have been observed to have certain regularities in statistical properties, including leptokurtic distributions, volatility clustering, leverage effects and persistence of volatility. Since the seminal work of Bollerslev (1986; 1990), the most popular approach uses generalized autoregressive conditional heteroskedastic (GARCH) models to capture many of these stylized facts in the data. The variance in such models is taken to be a function of past observations and variances. For instance, Engle (1982) estimates the variance of the UK inflation by GARCH. Theoretical results on the formulation and estimation of multivariate GARCH models within simultaneous equation systems can be found in Engle and Kroner (1995). Successively, Engle (2002) introduces a simple class of multivariate GARCH models, called dynamic conditional correlation (DCC) GARCH. These are not linear but have the flexibility of univariate GARCH coupled with parsimonious parametric models for the correlations, and provide sensible empirical results in the investigation of financial markets. For information concerning quasi-maximum likelihood estimation, strict stationarity, geometric ergodicity and asymptotic theory of multivariate GARCH models see Ling and McAleer (2003), Hafner and Preminger (2009), Boussama *et al.* (2011), and Francq and Zakoïan (2012). Conditions for the existence of fourth moments for such models and results for the kurtosis and cokurtosis between components have been derived by Hafner (2003). Recently, in the context of estimation for such models, Xuan *et al.* (2024) derive stochastic variational inference algorithms for fitting various GARCH-type models. Moreover, a multivariate heavy-tailed integer-valued GARCH process with EM algorithm-based inference has been proposed by Jang *et al.* (2024). Finally, Livingston Jr. and Nur (2023) present a Bayesian analysis on the BEKK formulation of the multivariate GARCH model by combining a set of existing MCMC algorithms in the literature.

✉ Maddalena Cavicchioli
maddalena.cavicchioli@unimore.it

Jie Cheng
j.cheng@keele.ac.uk

1 Department of Economics Marco Biagi, University of Modena and Reggio Emilia, Viale Berengario 51, 41121 Modena, Italy

2 School of Computer Science and Mathematics, Keele University, Keele, UK

However, economic and financial systems often suffer from shocks that shift them from their present state into another one. These states tend to be stochastic and dynamic. To capture such probabilistic state transitions over time, Markov switching (MS) models provide a useful framework for applied empirical work since the seminal papers of Hamilton (1989, 1990). In such models the parameters and the return distribution at a given time depend on the state (or regime) of an unobserved Markov chain. For example, Guidolin et al. (2012) show that simple VARs cannot approximate MS asset allocation decisions. Foroni *et al.* (2024) develop an expectile MS model for the analysis of cryptocurrency time series in a risk management framework.

Existing literature on *univariate* MS GARCH models includes studies focusing on estimation, autocovariance structure, statistical inference and spectral analysis. Such models are gradually becoming a useful tool for economist and financial practitioners for analyzing financial market data. In this setting we only mention some papers relate to our topic. Francq and Zakoïan (2008) propose a procedure for computing the autocovariances and the ARMA representations of the squares and higher-order powers of univariate MS GARCH models. Explicit derivation of the autocovariances allows for parameter estimation via GMM procedure. Augustyniak (2014) develops an approach based on both the Monte Carlo expectation-maximization algorithm and importance sampling to calculate the maximum likelihood (ML) estimator and its asymptotic variance of an univariate MS GARCH model. Higher order moments, kurtosis measures and spectral analysis of univariate MS GARCH models with statistical inference can be found in Cavicchioli (2022, 2021). For a general theory and inference for univariate MS GARCH models see Francq and Zakoïan (2005), Liu (2006), and Bauwens *et al.* 2010. A comprehensive approach for stationarity analysis of such models has been developed by Abramson and Cohen (2007). Hu and Shin (2008) proposed an optimal specification test against an MS GARCH process based on the LR principle.

More recently, to account for structural linkages and interdependencies among various financial variables, researchers have developed *multivariate* GARCH-type models with regime dependent population parameters. These specifications have proved particularly useful for analyzing volatility dynamics, correlation shifts, and risk management in complex markets. Among the foundational contributions, Engle (2002) introduced the dynamic conditional correlation (DCC) GARCH model, which was later extended by Billio and Caporin (2005) to include Markov regime switches in both the unconditional correlation matrix and the DCC parameters. This extension is motivated by the need to capture abrupt structural changes in financial systems, a feature that is essential for contagion analysis with critical implications for monetary and macroprudential policies.

Building on this line of research, several authors have proposed alternative multivariate volatility frameworks incorporating mixture or regime switching mechanisms. Bauwens *et al.* 2007, for example, model the conditional distribution of a vector time series as a mixture of multivariate normal distributions, each with a time-varying covariance matrix. Regime switching dynamic correlation models have been successfully applied to study financial contagion in stock markets (see Nielsen Rotta and Valls Pereira, 2016). Advances in Bayesian inference for multivariate MS GARCH models have been developed by Billio *et al.* 2016 and AusÃƒÂn and Galeano (2007), who propose efficient simulation techniques for Markov switching and mixture GARCH frameworks, respectively. Applications of vector MS GARCH models to study volatility spillovers and contagion effects, such as those observed between energy and financial sectors during the COVID-19 crisis, can be found in Ghorbel and Jeribi (2021). Theoretical developments, such as sufficient conditions for the existence of second and fourth moments in multivariate MS GARCH models, have been presented in Cavicchioli (2021), providing neat matrix expressions in closed form for such moments.

Recent years have also witnessed a surge in applications of MS GARCH models to the cryptocurrency markets, where volatility and correlation structures are highly non-stationary. Relevant contributions include Ardia *et al.* (2019), Jha and Baur (2020), Chkili (2021), and Tan *et al.* (2021). In addition, vector MS (BEKK) GARCH models are widely used in estimating time-varying minimum variance hedge ratios and examining the effects of asymmetries and regime switching on optimal futures hedging (see Lee and Yoder, 2007; Lee, 2008; Su and Wu, 2014; Zhiping and Shenghong, 2018). Other applications of multivariate MS GARCH models in the analysis of financial asset returns, stock price volatility, and exchange rate dynamics are provided in Alexander and Lazar (2009), Bu *et al.* (2017), and Haas and Liu (2018). Furthermore, risk measurement applications using MS GARCH models have been studied by Sajiad *et al.* (2008), Zhu and Cheng (2013), and Ardia *et al.* 2018. For comprehensive surveys of MS models in economics and finance, see Guidolin (2011), Ang and Timmermann (2012), and Hamilton (2016).

Despite the growing literature on multivariate MS GAR CH frameworks, the efficient estimation of Markov switching BEKK models remains relatively underexplored. Existing approaches often rely on computationally intensive Bayesian techniques, which can present substantial challenges in high-dimensional settings and may lack transparency in model updates. The present study aims to address this gap by introducing two novel proposals for estimating MS BEKK models using (extended) Kalman filter techniques. These proposals are based on distinct state space representations of the underlying models.

In statistics and econometrics, a filter is an algorithm for recursive estimation of unobserved, time-varying parameters or latent states in dynamic systems. In particular, the Kalman filter uses the current observation to predict the next period's value of the latent variables and subsequently updates these predictions based on realized observations. For a survey of Kalman filtering techniques in econometric modeling, see Pasricha (2006). To our knowledge, no systematic investigation exists on Kalman-filter-based estimation of MS BEKK models, which could offer an appealing alternative to Bayesian simulation methods by enhancing computational efficiency and facilitating online updating of parameter estimates.

The contributions of this paper are threefold. First, we develop two alternative filtering algorithms for MS BEKK models, providing a unified framework for recursive estimation and inference in regime switching multivariate volatility systems. Second, we perform extensive numerical simulations to assess the performance of the proposed algorithms under various data-generating processes, focusing on their accuracy, convergence properties, and computational feasibility. Third, we apply our methods to real-world financial data to investigate contagion effects in stock markets and compare the predictive performance of our approach with existing MS GARCH models, particularly in value-at-risk (VaR) forecasting. The proposed framework aims to provide practitioners and policymakers with more accessible tools for analyzing regime dependent volatility dynamics in high-dimensional financial environments.

The paper is organized as follows. Section 2 provides the description of the standard vector BEEK model from Engle and Kroner (1995) and its VARMA representation. Then such a model is generalized to a vector MS BEEK model in which the parameters are driven by an unobserved Markov switching process. Then we derive a MS VARMA representation and different state space descriptions of the proposed model. These serve to develop a theory of linear filtering for multivariate MS BEEK models. Because of path dependence, ML estimation of MS GARCH-type models is intractable. This occurs because the conditional variance at time t depends on the entire sequence of regimes up to time t due to the recursive nature of the GARCH process and state dependent GARCH coefficients. To circumvents the problem of path dependence, Section 3 proposes two recursive algorithms for filtering and estimation of vector MS BEEK model. Sections 4 and 5 are devoted to illustrate the usefulness of the proposed methodologies via several simulation experiments and empirical applications to financial real data, respectively. Finally, Section 6 concludes.

2 Multivariate Markov switching BEKK models

2.1 The standard BEKK model

The standard multivariate GARCH models are in general computationally demanding for the large number of parameters. So, in the literature, some authors have introduced more simple representations to reduce their computational burden. Among other, we mention the diagonal Vech-GARCH and the BEKK representations from Engle and Kroner (1995), and the constant conditional correlation (CCC) GARCH from Bollerslev (1990).

The standard n-dimensional BEKK(p, q) model is defined as

$$X_t = \Sigma_t^{1/2}\epsilon_t \tag{1}$$

$$\Sigma_t = CC' + \sum_{i=1}^{q}\sum_{k=1}^{l_i} A_{i,k}X_{t-i}X'_{t-i}A'_{i,k} + \sum_{i=1}^{p}\sum_{k=1}^{m_i} B_{i,k}\Sigma_{t-i}B'_{i,k} \tag{2}$$

The model automatically ensures positive semi-definiteness of $n \times n$ conditional covariances.

We restrict ourselves to the model with each l_i and m_i equal to one

$$X_t = \Sigma_t^{1/2}\epsilon_t \tag{3}$$

$$\Sigma_t = CC' + \sum_{i=1}^{q} A_i X_{t-i}X'_{t-i}A'_i + \sum_{i=1}^{p} B_i\Sigma_{t-i}B'_i \tag{4}$$

If we define $\varsigma_t = \text{vech}(\Sigma_t)$ and $y_t = \text{vech}\left(X_tX'_t\right)$, then equation (4) can be rewritten as

$$\varsigma_t = \tilde{C} + \sum_{i=1}^{q}\tilde{A}_i\, y_{t-i} + \sum_{i=1}^{p}\tilde{B}_i\,\varsigma_{t-i} \tag{5}$$

where

$$\tilde{C} = \text{vech}\left(CC'\right) \tag{6}$$

$$\tilde{A}_i = H_n\left(A_i \otimes A_i\right)K'_n \tag{7}$$

$$\tilde{B}_i = H_n\left(B_i \otimes B_i\right)K'_n \tag{8}$$

$$\text{vech}(M) = H_n\text{vec}(M) \tag{9}$$

$$\text{vec}(M) = K'_n\text{vech}(M) \tag{10}$$

$$H_nK'_n = I_{n(n+1)/2}. \tag{11}$$

The above notations are clarified as follows. The vec operator transforms a $(n \times n)$ matrix M into a $(n^2 \times 1)$ vector by stacking the columns of M one underneath the other. The vech operator stacks the lower triangular elements of a symmetric matrix. Here $K_n^{'}$ is the transpose matrix of the usual $m \times n^2$ *duplication matrix* D_n with $m = [n(n+1)]/2$, as defined, for example, in Magnus and Neudecker (1979, 1986). For symmetric M, the matrix $K_n^{'}$ transforms $\text{vech}(M)$ into $\text{vec}(M)$. Furthermore, H_n denotes the Moore-Penrose inverse of $K_n^{'}$, that is $H_n = (K_n K_n^{'})^{-1} K_n$ (denoted by D_n^{+} in the cited papers). Then, for symmetric M, H_n transforms $\text{vec}(M)$ into $\text{vech}(M)$. As usual, $\mathbb{R}^{m \times n}$ is the class of real $(m \times n)$ matrices, and $\mathbb{R}^n$ the class of real $(n \times 1)$ vectors, so that $\mathbb{R}^n \equiv \mathbb{R}^{n \times 1}$.

Setting $\eta_t = y_t - \varsigma_t$, we get the following VARMA(r, p) representation of the above BEKK(p, q) model

$$y_t = \tilde{C} + \sum_{i=1}^{r} \Phi_i \, y_{t-i} + \eta_t - \sum_{i=1}^{p} \tilde{B}_i \, \eta_{t-i} \tag{12}$$

$$\Phi_i = \begin{cases} \tilde{A}_i + \tilde{B}_i & 1 \le i \le s \\ \tilde{A}_i \mathbb{I}_{p<q} + \tilde{B}_i \mathbb{I}_{p>q} & s+1 \le i \le r \end{cases} \tag{13}$$

Here we set $r = \max(p, q)$ and $s = \min(p, q)$. The interesting thing about this representation is that if the innovations ϵ_t are martingale difference sequences with zero mean and identity covariance matrix, then η_t are such that $E[\eta_t] = E\left[\eta_t \,|\mathcal{I}_{t-1}\right] = 0$. This implies that they are also martingale difference sequences. Sufficient condition for stationarity is that the spectral radius of $\sum_{i=1}^{r} \Phi_i$ is stricly less than one.

2.2 The Markov Switching vector *BEKK* model

An n-dimensional MS BEKK(p, q) model is defined as

$$X_t = \Sigma_t^{1/2}(\mathcal{I}_{t-1}, s_t)\, \epsilon_t \tag{14}$$

$$\Sigma_t(\mathcal{I}_{t-1}, s_t) = C_{s_t} C_{s_t}^{'} + \sum_{i=1}^{q} \sum_{k=1}^{l_i} A_{s_t,ik} X_{t-i} X_{t-i}^{'} A_{s_t,ik}^{'} + \sum_{i=1}^{p} \sum_{k=1}^{m_i} B_{s_t,ik} \Sigma_{t-i}(\mathcal{I}_{t-i-1}, s_{t-i}) B_{s_t,ik}^{'} \tag{15}$$

where $\mathcal{I}_t = \{X_t, X_{t-1}, \dots\}$ is the information set and $\epsilon_t \sim IID(0, I_n)$. We assume that (s_t) is a homogeneous irreducible and aperiodic (hence ergodic) Markov chain which takes on values in the set $\Xi = \{1, \dots, d\}$, stationary transition probabilities denoted by $p_{ij} = Pr(s_t = j | s_{t-1} = i)$ and unconditional (or steady state) probabilities $\pi_i = Pr(s_t = i)$ for every $i \in \Xi$. Let $\boldsymbol{P} = (p_{ij})$ denote the *transition probability matrix* and $\boldsymbol{\pi} = (\pi_1, \dots, \pi_d)^{'}$ the *stationary vector* of the chain. In addition, (s_t) is assumed to be independent of (ϵ_t). The model ensures positive definiteness of $n \times n$ conditional variances.

We restrict ourselves to the model with each l_i and m_i equal to one

$$X_t = \Sigma_t^{1/2}(\mathcal{I}_{t-1}, s_t)\, \epsilon_t \tag{16}$$

$$\Sigma_t(\mathcal{I}_{t-1}, s_t) = C_{s_t} C_{s_t}^{'} + \sum_{i=1}^{q} A_{s_t,i} X_{t-i} X_{t-i}^{'} A_{s_t,i}^{'} + \sum_{i=1}^{p} B_{s_t,i} \Sigma_{t-i}(\mathcal{I}_{t-i-1}, s_{t-i}) B_{s_t,i}^{'}. \tag{17}$$

If we define

$$\varsigma_t(\mathcal{I}_{t-1}, s_t) = \text{vech}\, \Sigma_t(\mathcal{I}_{t-1}, s_t)$$

and

$$y_t = \text{vech}(X_t X_t^{'}),$$

then equation (17) can be rewritten as

$$\varsigma_t(\mathcal{I}_{t-1}, s_t) = \tilde{C}_{s_t} + \sum_{i=1}^{q} \tilde{A}_{s_t,i} \, y_{t-i} + \sum_{i=1}^{p} \tilde{B}_{s_t,i} \, \varsigma_{t-i}(\mathcal{I}_{t-i-1}, s_{t-i}) \tag{18}$$

where

$$\tilde{C}_{s_t} = \text{vech}\left(C_{s_t} C_{s_t}^{'}\right) \tag{19}$$

$$\tilde{A}_{s_t,i} = H_n \left(A_{s_t,i} \otimes A_{s_t,i}\right) K_n^{'} \tag{20}$$

$$\tilde{B}_{s_t,i} = H_n \left(B_{s_t,i} \otimes B_{s_t,i}\right) K_n^{'} \tag{21}$$

where H_n and $K_n^{'}$ are as in Equations $(9 - 11)$.

In order to develop a theory of linear filtering for multivariate MS BEKK models, we need to link the model with some state space representations. Here we propose a state space representation and write the associated filter. For this, we first derive a MS VARMA representation of the initial model (16–18).

Let us define

$$\eta_t(\mathcal{I}_{t-1}, s_t) = y_t - \varsigma_t(\mathcal{I}_{t-1}, s_t). \tag{22}$$

Then (η_t) is zero mean serially uncorrelated. However, (η_t) is not independent over time since it does not have a constant variance in time (i.e., it is not a homoskedastic process). By (5) we have

$$\begin{aligned} y_t = \varsigma_t + \eta_t &= \tilde{C}_{s_t} + \sum_{i=1}^{q} \tilde{A}_{s_t,i} \, y_{t-i} + \sum_{i=1}^{p} \tilde{B}_{s_t,i} \, \varsigma_{t-i} + \eta_t \\ &= \tilde{C}_{s_t} + \sum_{i=1}^{q} \tilde{A}_{s_t,i} \, y_{t-i} \\ &\quad + \sum_{i=1}^{p} \tilde{B}_{s_t,i} \, (y_{t-i} - \eta_{t-i}) + \eta_t \end{aligned}$$

$$= \tilde{C}_{s_t} + \sum_{i=1}^{r} \Phi_{s_t,i}\, y_{t-i} + \eta_t - \sum_{i=1}^{p} \tilde{B}_{s_t,i}\, \eta_{t-i} \tag{23}$$

where

$$\Phi_{s_t,i} = \begin{cases} \tilde{A}_{s_t,i} + \tilde{B}_{s_t,i} & 1 \le i \le s \\ \tilde{A}_{s_t,i}\mathbb{I}_{p<q} + \tilde{B}_{s_t,i}\mathbb{I}_{p>q} & s+1 \le i \le r \end{cases} \tag{24}$$

As before, we set $r = \max(p, q)$ and $s = \min(p, q)$. Equation (23) gives a MS VARMA(r, p) representation of the vector MS BEKK(p, q) model. The interesting thing about this representation is that if the innovations ϵ_t are martingale difference sequences with zero mean and identity covariance matrix, then η_t are such that $E(\eta_t) = E(\eta_t|\mathcal{I}_{t-1}) = 0$. This implies that they are also mantingale difference sequences. Sufficient condition for stationarity of such MS VARMA model is given in Francq and Zakoïan (2001). Results on the stationarity of MS GARCH models can be found in Francq and Zakoïan (2005), Liu (2006), and Abramson and Cohen (2007).

3 Filtering and estimation

We now propose two alternative filters for the estimation of the MS BEKK model defined in (14)–(15). In the following, we recast the model in a state space form and discuss the associated inference methods.

Distributional assumption (Gaussian working likelihood). Throughout this section we adopt a *Gaussian working likelihood* for the measurement innovation. Recall that

$$y_t = \operatorname{vech}(X_t X_t'), \qquad \eta_t = y_t - \varsigma_t(\mathcal{I}_{t-1}, s_t),$$

where $\varsigma_t(\mathcal{I}_{t-1}, s_t) = E[y_t \mid \mathcal{I}_{t-1}, s_t]$. Even if X_t is conditionally Gaussian, y_t is a quadratic form and therefore has a Wishart-/χ^2-type law. Consequently, neither y_t nor η_t is exactly Gaussian in finite samples. We nevertheless approximate the conditional law of η_t by

$$\eta_t \mid \mathcal{I}_{t-1}, s_t \overset{\text{approx}}{\sim} \mathcal{N}_m(\mathbf{0}, \mathbf{H}_{s_t}),$$

where $\mathbf{H}_{s_t}$ is the regime-specific covariance implied by the BEKK dynamics. This is *not* an assumption of normality for y_t; rather, it provides a tractable quasi-likelihood for the *centered* quadratic measurement error. The approximation is motivated by a CLT for centered quadratic forms (chi-square/Wishart asymptotics) when y_t aggregates many underlying shocks. See Appendix C of Supplementary Materials for a formal justification and the relevant normalization. Operationally, this choice places the model within the class of conditionally linear Gaussian state space systems *within each regime*, so that Kalman-type recursions and Kim filtering can be applied. Importantly, using a Gaussian likelihood for non-Gaussian disturbances is standard in state space inference as a quasi-maximum likelihood device. See Hamilton (1994a) [Ch. 13, § 13.4, *Quasi-Maximum Likelihood Estimation*], as well as [Ch. 5, § 5.3] for the general QML interpretation under misspecified Gaussian likelihoods. For Markov switching models and multi-move state simulation based on filtering/smoothing under conditionally Gaussian likelihoods, see Frühwirth-Schnatter (2006) [Ch. 11, § 11.2–11.3 and Subsec. 11.5.3] and Kim (1994).

3.1 First proposal

From (23) we get

$$y_t = \tilde{C}_{s_t} + \sum_{i=1}^{r} (y_{t-i}' \otimes I_m) \operatorname{vec} \Phi_{s_t,i}$$
$$+ \eta_t - \sum_{i=1}^{p} (\eta_{t-i}' \otimes I_m) \operatorname{vec} \tilde{B}_{s_t,i},$$

where $m = [n(n+1)]/2$. Define

$$X_t' = (I_m \quad y_{t-1}' \otimes I_m \quad \cdots \quad y_{t-r}' \otimes I_m \quad -\eta_{t-1}' \otimes I_m \quad \cdots \quad -\eta_{t-p}' \otimes I_m) \in \mathbb{R}^{m \times K}$$

hence $X_t \in \mathbb{R}^{K \times m}$, where $K = m + (p + r)m^2$, and

$$\boldsymbol{\beta}_{s_t} = \left(\tilde{C}_{s_t}' \quad [\operatorname{vec} \Phi_{s_t,1}]' \quad \cdots \quad [\operatorname{vec} \Phi_{s_t,r}]' \quad [\operatorname{vec} \tilde{B}_{s_t,1}]' \quad \cdots \quad [\operatorname{vec} \tilde{B}_{s_t,p}]' \right)' \in \mathbb{R}^K.$$

Let $\boldsymbol{\beta}_i$ denote the $(K \times 1)$-dimensional vector obtained from $\boldsymbol{\beta}_{s_t}$ by setting $s_t = i$, for every $i = 1, \ldots, d$. Set

$$\boldsymbol{B} = [\boldsymbol{\beta}_1 \cdots \boldsymbol{\beta}_d] \in \mathbb{R}^{K \times d}.$$

Let $\boldsymbol{\xi}_t$ be the $(d \times 1)$-dimensional vector whose ith component equals 1 if $s_t = i$ and zero otherwise. It is known that the Markov chain (s_t) admits the VAR(1) representation

$$\boldsymbol{\xi}_t = \boldsymbol{P}' \boldsymbol{\xi}_{t-1} + \boldsymbol{v}_t, \tag{25}$$

where $\boldsymbol{P} = (p_{ij})$ is the transition probability matrix, and $\boldsymbol{v}_t$ is a zero-mean innovation satisfying $E[\boldsymbol{v}_t \mid \mathcal{G}_{t-1}] = \mathbf{0}$. Here

$$\mathcal{F}_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\}$$

is the observable information set up to time $t-1$, and we define the enlarged information set

$$\mathcal{G}_{t-1} := \sigma(\mathcal{F}_{t-1}, s_{t-1}),$$

which additionally contains the regime at time $t-1$ (so that s_{t-1} is known conditional on $\mathcal{G}_{t-1}$). See Hamilton (1994a, §22) and Krolzig (1997, §2).

To specify the covariance structure of $\boldsymbol{v}_t$, observe that

$$\operatorname{Var}(\boldsymbol{v}_t \mid \mathcal{G}_{t-1}) = \operatorname{Var}(\boldsymbol{\xi}_t \mid \mathcal{G}_{t-1}) = E[\boldsymbol{\xi}_t\boldsymbol{\xi}_t' \mid \mathcal{G}_{t-1}] \\ -E[\boldsymbol{\xi}_t \mid \mathcal{G}_{t-1}]\, E[\boldsymbol{\xi}_t \mid \mathcal{G}_{t-1}]'.$$

The conditional expectation of $\boldsymbol{\xi}_t$ is $E[\boldsymbol{\xi}_t \mid \mathcal{G}_{t-1}] = \boldsymbol{P}'\boldsymbol{\xi}_{t-1}$. We clarify how this result is obtained. By the above definition of $\boldsymbol{\xi}_t$, it follows that

$$(\boldsymbol{\xi}_t)_i = \mathbb{I}(s_t = i),$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. Given the Markov property of (s_t), we have for any $i, j \in \{1, \ldots, d\}$,

$$\Pr(s_t = i \mid s_{t-1} = j, \mathcal{G}_{t-1}) = \Pr(s_t = i \mid s_{t-1} = j) = p_{ji},$$

where p_{ji} denotes the (j, i) entry of the transition matrix $\boldsymbol{P}$. Conditional on $\mathcal{G}_{t-1}$, the state s_{t-1} is known, hence $\boldsymbol{\xi}_{t-1}$ is non-stochastic. Therefore, for each component i of $\boldsymbol{\xi}_t$, we have

$$E[(\boldsymbol{\xi}_t)_i \mid \mathcal{G}_{t-1}] = \sum_{j=1}^{d} \Pr(s_t = i, s_{t-1} = j \mid \mathcal{G}_{t-1}) \\ = \sum_{j=1}^{d} \Pr(s_t = i \mid s_{t-1} \\ = j, \mathcal{G}_{t-1})\, \Pr(s_{t-1} = j \mid \mathcal{G}_{t-1}) \\ = \sum_{j=1}^{d} p_{ji}\, (\boldsymbol{\xi}_{t-1})_j .$$

Stacking these d equations in vector form yields $E[\boldsymbol{\xi}_t \mid \mathcal{G}_{t-1}] = \boldsymbol{P}'\boldsymbol{\xi}_{t-1}$, as claimed. Moreover, the second moment of $\boldsymbol{\xi}_t$ is $E[\boldsymbol{\xi}_t\boldsymbol{\xi}_t' \mid \mathcal{G}_{t-1}] = \operatorname{diag}(\boldsymbol{P}'\boldsymbol{\xi}_{t-1})$. Therefore, the conditional covariance matrix of $\boldsymbol{v}_t$ is given by

$$\operatorname{Var}(\boldsymbol{v}_t \mid \mathcal{G}_{t-1}) = \operatorname{diag}(\boldsymbol{P}'\boldsymbol{\xi}_{t-1}) - (\boldsymbol{P}'\boldsymbol{\xi}_{t-1})(\boldsymbol{P}'\boldsymbol{\xi}_{t-1})'.$$

In the stationary case, if $\boldsymbol{\pi}$ is the invariant distribution such that $\boldsymbol{P}'\boldsymbol{\pi} = \boldsymbol{\pi}$, then the unconditional covariance matrix of $\boldsymbol{v}_t$ is

$$\operatorname{Var}(\boldsymbol{v}_t) = \sum_{i=1}^{d} \pi_i \left(\operatorname{diag}(\boldsymbol{p}_i) - \boldsymbol{p}_i \boldsymbol{p}_i'\right),$$

where $\boldsymbol{p}_i$ denotes the i-th column of $\boldsymbol{P}'$, that is, the transition probabilities from state i.

If one conditions only on the observable information $\mathcal{F}_{t-1}$, then s_{t-1} is typically latent and $\boldsymbol{\xi}_{t-1}$ is not $\mathcal{F}_{t-1}$-measurable. Define the filtered regime-probability vector

$$\boldsymbol{\xi}_{t-1|t-1} := E[\boldsymbol{\xi}_{t-1} \mid \mathcal{F}_{t-1}] \\ = \left(\Pr(s_{t-1} = 1 \mid \mathcal{F}_{t-1}), \ldots, \Pr(s_{t-1} = d \mid \mathcal{F}_{t-1})\right)'.$$

Then iterated expectations give

$$E[\boldsymbol{\xi}_t \mid \mathcal{F}_{t-1}] = E\big[E[\boldsymbol{\xi}_t \mid \mathcal{G}_{t-1}] \mid \mathcal{F}_{t-1}\big] = \boldsymbol{P}'\, \boldsymbol{\xi}_{t-1|t-1}.$$

In the stationary case, writing $\boldsymbol{p}_j := (p_{j1}, \ldots, p_{jd})'$ for the transition-probability vector out of state j, the above expression for $\operatorname{Var}(\boldsymbol{v}_t)$ holds too.

Then (23) admits the state space representation

$$\begin{cases} y_t = \boldsymbol{X}_t'\, \boldsymbol{B}\, \boldsymbol{\xi}_t + \eta_t \\ \boldsymbol{\xi}_t = \boldsymbol{P}'\, \boldsymbol{\xi}_{t-1} + \boldsymbol{v}_t. \end{cases} \tag{26}$$

Note that $\boldsymbol{B}\,\boldsymbol{\xi}_t = \boldsymbol{\beta}_{s_t}$, hence $\boldsymbol{X}_t'\,\boldsymbol{B}\,\boldsymbol{\xi}_t = \boldsymbol{X}_t'\,\boldsymbol{\beta}_{s_t}$. To compute $\boldsymbol{X}_t'$ we can write

$$\operatorname{vec} \varsigma_t = \tilde{C}_{s_t} + \sum_{i=1}^{q} (y_{t-i}' \otimes I_m)\, \operatorname{vec} \tilde{A}_{s_t,i} \\ + \sum_{i=1}^{p} (\varsigma_{t-i}' \otimes I_m)\, \operatorname{vec} \tilde{B}_{s_t,i}.$$

by using (18).

Define

$$\boldsymbol{Z}_t' = (I_m \quad y_{t-1}' \otimes I_m \; \cdots \; y_{t-q}' \otimes I_m \quad \varsigma_{t-1}' \otimes I_m \; \cdots \\ \varsigma_{t-p}' \otimes I_m) \in \mathbb{R}^{m \times L}$$

where $L = m + (p+q)m^2$. Then we have

$$\boldsymbol{\lambda}_{s_t} = \left(\tilde{C}_{s_t}' \quad [\operatorname{vec} \tilde{A}_{s_t,1}]' \quad \cdots \quad [\operatorname{vec} \tilde{A}_{s_t,q}]' \quad [\operatorname{vec} \tilde{B}_{s_t,1}]' \right. \\ \left. \cdots \quad [\operatorname{vec} \tilde{B}_{s_t,p}]'\right)' \in \mathbb{R}^L$$

and $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1 \quad \cdots \quad \boldsymbol{\lambda}_d) \in \mathbb{R}^{L \times d}$. It follows the recursive formula

$$\varsigma_t = \boldsymbol{Z}_t' \boldsymbol{\Lambda} \boldsymbol{\xi}_t \qquad \text{and} \qquad \eta_t = y_t - \varsigma_t \tag{27}$$

Given $\varsigma_{-1}, \ldots, \varsigma_{-p}$ and $y_{-1}, \ldots, y_{-q}$, we have $\boldsymbol{Z}_0'$. Then we compute ς_0 by (27) and so on iterating.

Remark (Gaussian working likelihood for the measurement innovation and Kalman filtering) Equation (26) rewrites the multivariate Markov switching BEKK dynamics in state space form, where the latent regime process (s_t) is encoded by the d-dimensional indicator vector $\boldsymbol{\xi}_t$ satisfying the VAR(1) representation in (25), while the observed process y_t is governed by the measurement equation in the first row of (26). Filtering and likelihood evaluation in such models rely crucially on distributional assumptions for the measurement innovation η_t and the state innovation $\boldsymbol{v}_t$.

(a) The exact law of y_t is non-Gaussian. By construction $y_t = \text{vech}(X_t X_t')$ is a vector of quadratic forms in the return vector X_t. Even when $X_t \,|_{\mathcal{I}_{t-1},s_t}$ is conditionally Gaussian with conditional covariance $\Sigma_t(s_t)$, the outer product $X_t X_t'$ is (up to scaling) Wishart with one degree of freedom, and each component of y_t has a χ^2-type (skewed, bounded-below) distribution. Consequently, the centered innovation

$$\eta_t \;=\; y_t - \varsigma_t(\mathcal{I}_{t-1}, s_t), \qquad \varsigma_t(\mathcal{I}_{t-1}, s_t) = E\big[y_t \mid \mathcal{I}_{t-1}, s_t\big],$$

is *not* exactly Gaussian at finite sample sizes; rather it is a centered (shifted) χ^2/Wishart-type object, hence generally asymmetric and bounded below by $-\varsigma_t(\mathcal{I}_{t-1}, s_t)$.

(b) Why a Gaussian *working* likelihood is nevertheless appropriate here. In our framework, Gaussianity of η_t is used as a *working likelihood* that delivers a tractable filter and likelihood for the Markov switching BEKK dynamics. This choice is supported by two complementary arguments. (i) *CLT-based Gaussian approximation of quadratic measurements.* In many financial applications the quadratic measurement $X_t X_t'$ is effectively an aggregation of many micro-level contributions (e.g., sums of intraday outer products, or returns built from many micro shocks). In such cases the relevant quadratic statistic is close to a (scaled) Wishart matrix with growing degrees of freedom, and the corresponding centered and suitably standardized vech$(\cdot)$ admits a multivariate Gaussian approximation via chi-square/Wishart CLT asymptotics. A precise statement of this approximation and its conditions is provided in the Supplementary Materials (Part C). (ii) *Quasi-Gaussian filtering under non-normality.* Even when the true measurement distribution is non-Gaussian, the Kalman recursions retain a clear optimality interpretation as the best linear least-squares predictor among all filters based on linear functions of the information set; see Hamilton (1994b, Chapter 50, § 2.5)[1]. Accordingly, we assume the conditional Gaussian working model

$$\eta_t \,|_{\mathcal{I}_{t-1},s_t} \approx \mathcal{N}_m\big(\mathbf{0}, \boldsymbol{H}_{s_t}\big),$$

where $\boldsymbol{H}_{s_t}$ is interpreted as the conditional second-moment matrix of the centered quadratic measurement under regime s_t (and, in our specification, is parameterized through the BEKK dynamics). Under this working likelihood, the resulting filter is computationally efficient and, by construction, targets the correct first and second conditional moments.

(c) The innovation $\boldsymbol{v}_t$ and switching-state inference. Recall that $\boldsymbol{\xi}_t \in \{e_1, \ldots, e_d\}$ almost surely; the VAR(1) representation in (25) is an algebraic encoding of the discrete Markov chain, and $\boldsymbol{v}_t$ represents regime jumps. Hence $\boldsymbol{v}_t$ is neither Gaussian nor continuous in the classical sense. Its role in the state space formulation is purely to represent the stochastic transition between regimes.

(d) Filtering strategy. Because the model is linear-Gaussian *conditional* on the regime sequence (s_t), inference is carried out within the conditionally linear Gaussian state space framework. We combine Kalman filtering for the continuous component conditional on s_t with the forward-backward recursion for the discrete latent Markov chain (Hamilton/Kim filtering). These hybrid filters are specifically designed for Markov switching state space models and are now standard tools in econometric and statistical applications.

(e) Applications. Empirically, the adequacy of the Gaussian assumption for η_t can be verified through residual diagnostics and normality tests within each regime. In our real-data application (see §5), the corresponding residual analysis is reported in Figures B.1 and B.2 of Supplementary Materials (B. RESIDUAL ANALYSIS), where for each component (FTSE100 and S&P500) and each regime, we present the standardized residual series together with the associated Q–Q plots, histograms, and autocorrelation functions. Across all cases, the residual diagnostics indicate that the dynamic specification of the model is adequate. In particular, the autocorrelation functions show no significant serial dependence in either regime, and the Q–Q plots exhibit reasonable alignment with Gaussian quantiles over most of the support.

The histograms display right-skewed distribution with strictly positive support, this pattern is theoretically well justified. As emphasised in the article, each component of η_t can be viewed as a quadratic form in Gaussian innovations and therefore, in finite samples, naturally follows χ^2-type distributions rather than a Gaussian one. Such χ^2-type behaviour is not surprising and is fully consistent with the diagnostics. Moreover, Monte Carlo simulations (see §4) confirm that the regime-specific Gaussian approximation still yields reliable inference for both the model parameters and the latent states, even in moderately sized samples. All simulations

[1] The specific discussion is in Hamilton (1994b, Chapter 50 "State-Space Models", §2.5 "Interpretation of the Kalman filter with non-normal disturbances") and in the regime-switching part of the same chapter (§4.1–§4.2). For switching linear Gaussian state space models and approximate filtering (Kim-type collapsing), see FrÃƒÂ¼hwirth-Schnatter (2006, Chapter 13), in particular §§13.1.2–13.1.4 and §13.3.5.

were carried out by generating data from a multivariate Gaussian innovation process and estimating the model using the proposed methods. The resulting performance was satisfactory across all scenarios. This supports the robustness of the methodology in practical sample sizes.

In summary, while y_t in the first row of (26) is not exactly Gaussian, its conditional innovation η_t can be well approximated by a multivariate normal distribution. This approximation, together with the conditionally linear structure of the model, justifies the use of Kalman-type filtering methods within each regime. The resulting hybrid procedure - combining regime switching and Kalman filters - is therefore both theoretically sound and empirically robust. □

The transition equation in (26) differs from a stable linear VAR(1) process by the fact that one eigenvalue of $\boldsymbol{P}'$ is equal to one as $\boldsymbol{P}'\boldsymbol{\pi} = \boldsymbol{\pi}$. Furthermore, the covariance matrix is singular due to the adding-up restriction, which however ensures stability. For analytical purposes, a slightly different formulation of the transition equation is more useful, where the identity $\boldsymbol{i}_d'\boldsymbol{\xi}_t = 1$ is eliminated (here $\boldsymbol{i}_d$ denotes the $d \times 1$ vector of ones). This procedure alters the state space representation (26) and one has to consider a new state vector of dimension $d-1$. See Krolzig (1997, §5).
More precisely, define

$$\boldsymbol{F} = \begin{pmatrix} p_{11} - p_{d,1} & \cdots & p_{d-1,1} - p_{d,1} \\ \vdots & & \vdots \\ p_{1,d-1} - p_{d,d-1} & \cdots & p_{d-1,d-1} - p_{d,d-1} \end{pmatrix}.$$

Then $\rho(\boldsymbol{F}) < 1$, where $\rho(\cdot)$ denotes the spectral radius. Set

$$\boldsymbol{\delta}_t = \begin{pmatrix} \xi_{1,t} - \pi_1 & \cdots & \xi_{d-1,t} - \pi_{d-1} \end{pmatrix}' \in \mathbb{R}^{d-1}. \tag{28}$$

From (26) we get the unrestricted state space representation

$$\begin{cases} y_t = \boldsymbol{X}_t' \tilde{\boldsymbol{B}} \boldsymbol{\delta}_t + \boldsymbol{X}_t' \boldsymbol{B} \boldsymbol{\pi} + \eta_t \\ \boldsymbol{\delta}_t = \boldsymbol{F} \boldsymbol{\delta}_{t-1} + \boldsymbol{w}_t \end{cases} \tag{29}$$

where $E(\boldsymbol{w}_t \boldsymbol{w}_t') = \tilde{\boldsymbol{D}} - \boldsymbol{F}\tilde{\boldsymbol{D}}\boldsymbol{F}' = \boldsymbol{Q}$, $\tilde{\boldsymbol{B}} = [\boldsymbol{\beta}_1 - \boldsymbol{\beta}_d \quad \cdots \quad \boldsymbol{\beta}_{d-1} - \boldsymbol{\beta}_d]$, and

$$\tilde{\boldsymbol{D}} = \begin{pmatrix} \pi_1(1-\pi_1) & -\pi_1\pi_2 & \cdots & -\pi_1\pi_{d-1} \\ -\pi_1\pi_2 & \pi_2(1-\pi_2) & \cdots & -\pi_2\pi_{d-1} \\ \vdots & \vdots & & \vdots \\ -\pi_{d-1}\pi_1 & -\pi_{d-1}\pi_2 & \cdots & \pi_{d-1}(1-\pi_{d-1}) \end{pmatrix}.$$

To derive the covariance structure of $\boldsymbol{w}_t$, we are given the $(d-1) \times 1$ vector $\boldsymbol{\delta}_t$ in (28) and the $d \times 1$ vector $\boldsymbol{\xi}_t$ that satisfies the restriction $\boldsymbol{i}_d'\boldsymbol{\xi}_t = 1$ with $\boldsymbol{i}_d = (1, \ldots, 1)' \in \mathbb{R}^d$. Hence, $\boldsymbol{\xi}_t$ is a vector of proportions and lies in the $(d-1)$-dimensional simplex:

$$\Delta^{d-1} = \left\{ \boldsymbol{\xi} \in \mathbb{R}^d : \xi_i \geq 0,\ \sum_{i=1}^{d} \xi_i = 1 \right\}.$$

Let $\boldsymbol{\xi}_t \in \mathbb{R}^d$ follow the transition equation in (25), where $\boldsymbol{P}'\boldsymbol{\pi} = \boldsymbol{\pi}$, so that $\boldsymbol{\pi}$ is a fixed point of the transition matrix. Define the centered process

$$\boldsymbol{\zeta}_t = \boldsymbol{\xi}_t - \boldsymbol{\pi}, \quad \text{so that} \quad \boldsymbol{\zeta}_t = \boldsymbol{P}'\boldsymbol{\zeta}_{t-1} + \boldsymbol{v}_t.$$

Let us denote by $\boldsymbol{\Sigma}_\zeta = \mathrm{Var}(\boldsymbol{\zeta}_t)$ the stationary variance of $\boldsymbol{\zeta}_t$. Since $\boldsymbol{\xi}_t$ lies in the unit simplex, it is well known (e.g. from multinomial or Dirichlet models) that the covariance matrix $\boldsymbol{\Sigma}_\zeta$ is singular, and can be written as

$$\boldsymbol{\Sigma}_\zeta = \boldsymbol{D} = \mathrm{Var}(\boldsymbol{\xi}_t) = \mathrm{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}'.$$

To reduce dimensionality and eliminate the singularity, define $\boldsymbol{\delta}_t$ as in (28). This is done by removing the d-th component of $\boldsymbol{\zeta}_t$, since $\xi_{d,t} = 1 - \sum_{i=1}^{d-1} \xi_{i,t}$, and analogously for π_d. Hence, $\delta_{i,t} = \zeta_{i,t} = \xi_{i,t} - \pi_i$, for all $i = 1, \ldots, d-1$. Let $\boldsymbol{T}$ be the $(d-1) \times d$ matrix that extracts the first $d-1$ elements, i.e.,

$$\boldsymbol{T} = \begin{pmatrix} I_{d-1} & O_{(d-1)\times 1} \end{pmatrix},$$

so that $\boldsymbol{\delta}_t = \boldsymbol{T}\boldsymbol{\zeta}_t$. Here I_{d-1} is the $(d-1) \times (d-1)$ identity matrix and $O_{(d-1)\times 1}$ is the $(d-1) \times 1$ vector of zeros. Then the variance of $\boldsymbol{\delta}_t$ is given by

$$\mathrm{Var}(\boldsymbol{\delta}_t) = \boldsymbol{T}\boldsymbol{D}\boldsymbol{T}' =: \tilde{\boldsymbol{D}}.$$

Now, we consider the dynamics of $\boldsymbol{\delta}_t$ from (29), where $\boldsymbol{w}_t = \boldsymbol{\delta}_t - \boldsymbol{F}\boldsymbol{\delta}_{t-1}$ is a white noise process. Assuming that the process is covariance stationary, we compute

$$\mathrm{Var}(\boldsymbol{\delta}_t) = \boldsymbol{F}\, \mathrm{Var}(\boldsymbol{\delta}_{t-1}) \boldsymbol{F}' + \mathrm{Var}(\boldsymbol{w}_t).$$

By stationarity, it follows that $\mathrm{Var}(\boldsymbol{\delta}_t) = \tilde{\boldsymbol{D}}$, and therefore we get

$$\tilde{\boldsymbol{D}} = \boldsymbol{F}\tilde{\boldsymbol{D}}\boldsymbol{F}' + \mathrm{Var}(\boldsymbol{w}_t).$$

Solving for the innovation covariance yields

$$\mathrm{Var}(\boldsymbol{w}_t) = \tilde{\boldsymbol{D}} - \boldsymbol{F}\tilde{\boldsymbol{D}}\boldsymbol{F}' =: \boldsymbol{Q},$$

as stated after Equation (29).

Following notation from Kim and Nelson (1999), denote

$$
\begin{aligned}
\delta_{t|t-1}^{(i,j)} &= E[\boldsymbol{\delta}_t|\mathcal{F}_{t-1}, s_t = j, s_{t-1} = i] \\
\delta_{t|t-1} &= E[\boldsymbol{\delta}_t|\mathcal{F}_{t-1}] \\
\delta_{t-1|t-1}^{i} &= E[\boldsymbol{\delta}_{t-1}|\mathcal{F}_{t-1}, s_{t-1} = i] \\
P_{t|t-1}^{(i,j)} &= E[(\boldsymbol{\delta}_t - \delta_{t|t-1})(\boldsymbol{\delta}_t - \delta_{t|t-1})^{'}|\mathcal{F}_{t-1}, s_t \\
&= j, s_{t-1} = i] \\
P_{t-1|t-1}^{i} &= E[(\boldsymbol{\delta}_{t-1} - \delta_{t-1|t-1}^{i})(\boldsymbol{\delta}_{t-1} \\
&\quad - \delta_{t-1|t-1}^{i})^{'}|\mathcal{F}_{t-1}, s_{t-1} = i] \\
y_{t|t-1}^{(i,j)} &= E[y_t|\mathcal{F}_{t-1}, s_t = j, s_{t-1} = i] \\
\eta_{t|t-1}^{(i,j)} &= y_t - y_{t|t-1}^{(i,j)} \\
f_{t|t-1}^{(i,j)} &= E[\eta_{t|t-1}^{(i,j)}\eta_{t|t-1}^{(i,j)'}|\mathcal{F}_{t-1}, s_t = j, s_{t-1} = i].
\end{aligned}
$$

Conditional on $s_{t-1} = i$ and $s_t = j$ we obtain the following filter

Prediction

- $\delta_{t|t-1}^{(i,j)} = \boldsymbol{F}\,\delta_{t-1|t-1}^{i}$
- $P_{t|t-1}^{(i,j)} = \boldsymbol{F}\,P_{t-1|t-1}^{i}\,\boldsymbol{F}' + \boldsymbol{Q}$
- $\eta_{t|t-1}^{(i,j)} = y_t - \boldsymbol{X}_t'\,\tilde{\boldsymbol{B}}\,\delta_{t|t-1}^{(i,j)} - \boldsymbol{X}_t'\,\boldsymbol{B}\,\boldsymbol{\pi} = \boldsymbol{X}_t'\,\tilde{\boldsymbol{B}}\,(\delta_t - \delta_{t|t-1}^{(i,j)}) + \eta_t$
- $f_{t|t-1}^{(i,j)} = \boldsymbol{X}_t'\,\tilde{\boldsymbol{B}}\,P_{t|t-1}^{(i,j)}\,\tilde{\boldsymbol{B}}'\,\boldsymbol{X}_t + \boldsymbol{R}_j$

where

$$\boldsymbol{R}_j = E(\eta_t\eta_t'|\mathcal{F}_{t-1}, s_t = j) \in \mathbb{R}^{m\times m}.$$

Updating

- $\delta_{t|t}^{(i,j)} = \delta_{t|t-1}^{(i,j)} + K_t^{(i,j)}\,\eta_{t|t-1}^{(i,j)}$
- $P_{t|t}^{(i,j)} = (I_{d-1} - K_t^{(i,j)}\,\boldsymbol{X}_t'\,\tilde{\boldsymbol{B}})\,P_{t|t-1}^{(i,j)}$

where $K_t^{(i,j)} = P_{t|t-1}^{(i,j)}\,\tilde{\boldsymbol{B}}'\,\boldsymbol{X}_t\,[f_{t|t-1}^{(i,j)}]^{-1} \in \mathbb{R}^{(d-1)\times m}$ is the Kalman gain.

Initial conditions

- $\delta_{0|0}^{i} = E[\boldsymbol{\delta}_0|s_0 = i] = \mathbf{0}$
- $\mathrm{vec}(P_{0|0}^{i}) = (I_{(d-1)^2} - \boldsymbol{F}\otimes\boldsymbol{F})^{-1}\,\mathrm{vec}\,\boldsymbol{Q}$

The derivation of the filter (including notation, conditioning and clarifications) is detailed in the Supplementary Materials (A. KALMAN FILTER).

3.2 Second proposal

We start specifying some notations used through this subsection: I_n is the $n\times n$ identity matrix as usual, $O_{m\times n}$ denotes the matrix of order $m\times n$ whose entries are zeros. In particular, $O_{m\times 1}$ is the $m\times 1$ (column) vector of zeros and $O_{m\times 1}' = O_{1\times m}$. For simplicity, we set $O_n := O_{n\times n}$ and $O_n := 0$ (zero scalar) if $n = 1$. From (26) let us define

$$B_t = \left(y_t'\; y_{t-1}' \cdots y_{t-r+1}'\; \eta_t'\; \eta_{t-1}' \cdots \eta_{t-p+1}'\right)' \in \mathbb{R}^{\ell},$$

where $\ell = (p+r)m$ and $m = [n(n+1)]/2$. Then we have

$$B_t = \mu_{s_t} + F_{s_t}\,B_{t-1} + G\,\eta_t \tag{30}$$

where

$$
\begin{aligned}
\mu_{s_t} &= \left(\tilde{C}_{s_t}'\; O_{1\times m} \cdots O_{1\times m}\right)' \in \mathbb{R}^{\ell} \\
G &= \left(I_m\; O_m \cdots O_m\; I_m\; O_m \cdots O_m\right)' \in \mathbb{R}^{\ell\times m}
\end{aligned}
$$

and

$$
F_{s_t} = \begin{pmatrix}
\Phi_{s_t,1} & \Phi_{s_t,2} & \cdots & \Phi_{s_t,r-1} & \Phi_{s_t,r} & -\tilde{B}_{s_t,1} & \cdots & -\tilde{B}_{s_t,p-1} & -\tilde{B}_{s_t,p} \\
I_m & O_m & \cdots & O_m & O_m & O_m & \cdots & O_m & O_m \\
O_m & I_m & \cdots & O_m & O_m & O_m & \cdots & O_m & O_m \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots \\
O_m & O_m & \cdots & I_m & O_m & O_m & \cdots & O_m & O_m \\
O_m & O_m & \cdots & O_m & O_m & O_m & \cdots & O_m & O_m \\
O_m & O_m & \cdots & O_m & O_m & I_m & \cdots & O_m & O_m \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots \\
O_m & O_m & \cdots & O_m & O_m & O_m & \cdots & I_m & O_m
\end{pmatrix} \in \mathbb{R}^{\ell\times\ell}.
$$

Now we illustrate the use of G with three examples.

Example 1 For $r = p = 1$, we have

$$y_t = \tilde{C}_{s_t} + \Phi_{s_t,1}\,y_{t-1} + \eta_t - \tilde{B}_{s_t,1}\,\eta_{t-1}.$$

The matrix F_{s_t} is given by

$$F_{s_t} = \begin{pmatrix} \Phi_{s_t,1} & -\tilde{B}_{s_t,1} \\ O_m & O_m \end{pmatrix} \in \mathbb{R}^{(2m)\times(2m)}.$$

In fact, we get

$$B_t = \begin{pmatrix} y_t \\ \eta_t \end{pmatrix} = \begin{pmatrix} \tilde{C}_{s_t} \\ O_{m\times 1} \end{pmatrix} + \begin{pmatrix} \Phi_{s_t,1} & -\tilde{B}_{s_t,1} \\ O_m & O_m \end{pmatrix}\begin{pmatrix} y_{t-1} \\ \eta_{t-1} \end{pmatrix} + \begin{pmatrix} I_m \\ I_m \end{pmatrix}\eta_t$$

where $\mu_{s_t} = \left(\tilde{C}_{s_t}'\; O_{1\times m}\right)' \in \mathbb{R}^{(2m)\times 1}$ and $G = (I_m \quad I_m)' \in \mathbb{R}^{(2m)\times m}$.

Example 2 For $r = 2$ and $p = 1$, we have

$$y_t = \tilde{C}_{s_t} + \Phi_{s_t,1}\,y_{t-1} + \Phi_{s_t,2}\,y_{t-2} + \eta_t - \tilde{B}_{s_t,1}\,\eta_{t-1}.$$

The matrix F_{s_t} is given by

$$F_{s_t} = \begin{pmatrix} \Phi_{s_t,1} & \Phi_{s_t,2} & -\tilde{B}_{s_t,1} \\ I_m & O_m & O_m \\ O_m & O_m & O_m \end{pmatrix} \in \mathbb{R}^{(3m)\times(3m)}.$$

In fact, we get

$$B_t = \begin{pmatrix} y_t \\ y_{t-1} \\ \eta_t \end{pmatrix} = \begin{pmatrix} \tilde{C}_{s_t} \\ O_{m\times 1} \\ O_{m\times 1} \end{pmatrix} + \begin{pmatrix} \Phi_{s_t,1} & \Phi_{s_t,2} & -\tilde{B}_{s_t,1} \\ I_m & O_m & O_m \\ O_m & O_m & O_m \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \eta_{t-1} \end{pmatrix} + \begin{pmatrix} I_m \\ O_m \\ I_m \end{pmatrix} \eta_t$$

where $\mu_{s_t} = \left(\tilde{C}'_{s_t}\ O_{1\times m}\ O_{1\times m}\right)' \in \mathbb{R}^{(3m)\times 1}$ and $G = (I_m \quad O_m \quad I_m)' \in \mathbb{R}^{(3m)\times m}$.

Example 3 For $r = p = 2$, we have

$$y_t = \tilde{C}_{s_t} + \Phi_{s_t,1}\, y_{t-1} + \Phi_{s_t,2}\, y_{t-2} + \eta_t - \tilde{B}_{s_t,1}\, \eta_{t-1} - \tilde{B}_{s_t,2}\, \eta_{t-2}.$$

The matrix F_{s_t} is given by

$$F_{s_t} = \begin{pmatrix} \Phi_{s_t,1} & \Phi_{s_t,2} & -\tilde{B}_{s_t,1} & -\tilde{B}_{s_t,2} \\ I_m & O_m & O_m & O_m \\ O_m & O_m & O_m & O_m \\ O_m & O_m & I_m & O_m \end{pmatrix} \in \mathbb{R}^{(4m)\times(4m)}.$$

Then we get

$$B_t = \begin{pmatrix} y_t \\ y_{t-1} \\ \eta_t \\ \eta_{t-1} \end{pmatrix} = \begin{pmatrix} \tilde{C}_{s_t} \\ O_{m\times 1} \\ O_{m\times 1} \\ O_{m\times 1} \end{pmatrix} + \begin{pmatrix} \Phi_{s_t,1} & \Phi_{s_t,2} & -\tilde{B}_{s_t,1} & -\tilde{B}_{s_t,2} \\ I_m & O_m & O_m & O_m \\ O_m & O_m & O_m & O_m \\ O_m & O_m & I_m & O_m \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \eta_{t-1} \\ \eta_{t-2} \end{pmatrix} + \begin{pmatrix} I_m \\ O_m \\ I_m \\ O_m \end{pmatrix} \eta_t$$

where $\mu_{s_t} = \left(\tilde{C}'_{s_t}\ O_{1\times m}\ O_{1\times m}\ O_{1\times m}\right)' \in \mathbb{R}^{(4m)\times 1}$ and $G = (I_m \quad O_m \quad I_m \quad O_m)' \in \mathbb{R}^{(4m)\times m}$.

Therefore, for every s_t, we obtain the following MS state space representation of the multivariate MS BEKK(p, q) model in (16–18):

$$\begin{cases} y_t = H B_t \\ B_t = \mu_{s_t} + F_{s_t}\, B_{t-1} + G\, \eta_t \end{cases} \tag{31}$$

or equivalently

$$\begin{cases} y_t = H B_t \\ B_t = \boldsymbol{\mu}\, \boldsymbol{\xi}_t + \mathbb{F}\, (\boldsymbol{\xi}_t \otimes I_\ell)\, B_{t-1} + G\, \eta_t \\ \boldsymbol{\xi}_t = \boldsymbol{P}'\, \boldsymbol{\xi}_{t-1} + \boldsymbol{v}_t \end{cases}$$

where $H = \left(I_m\ O_m\ \cdots\ O_m\right) \in \mathbb{R}^{m\times\ell}$, $\boldsymbol{\mu} = (\mu_1\ \cdots\ \mu_d) \in \mathbb{R}^{\ell\times d}$, and $\mathbb{F} = (F_1\ \cdots\ F_d) \in \mathbb{R}^{\ell\times(\ell d)}$. Of course, we can also derive a unrestricted state space representation as done in (29). Representation (31) is similar but different to the switching dynamic model of Kim (1994) and Kim and Nelson (1999). The state vector B_t is called the state (of the system) at time t. According to Gourieroux and Monfort (1997, p.576), it can be partially unobservable as it is in this case.
Conditional on $s_{t-1} = k$ and $s_t = j$, we obtain the following filter (the derivation of this filtering procedure is similar to that detailed in the Supplementary Materials; therefore, it is omitted for brevity).

Prediction

- $B^{(k,j)}_{t|t-1} = \mu_j + F_j\, B^k_{t-1|t-1}$
- $P^{(k,j)}_{t|t-1} = F_j\, P^k_{t-1|t-1}\, F'_j + G\, Q_j\, G'$
- $\eta^{(k,j)}_{t|t-1} = y_t - y^{(k,j)}_{t|t-1} = y_t - H\, B^{(k,j)}_{t|t-1}$
- $f^{(k,j)}_{t|t-1} = H\, P^{(k,j)}_{t|t-1}\, H'$

Updating

- $B^{(k,j)}_{t|t} = B^{(k,j)}_{t|t-1} + K^{(k,j)}_t\, \eta^{(k,j)}_{t|t-1}$
- $P^{(k,j)}_{t|t} = P^{(k,j)}_{t|t-1} - K^{(k,j)}_t\, H\, P^{(k,j)}_{t|t-1}$

where $Q_j = \mathrm{Var}(\eta_t|\mathcal{F}_{t-1}, s_t = j) = \mathrm{Var}(\eta_t|s_t = j) \in \mathbb{R}^{m\times m}$ and $K^{(k,j)}_t = P^{(k,j)}_{t|t-1} H' [f^{(k,j)}_{t|t-1}]^{-1} \in \mathbb{R}^{\ell\times m}$ is the Kalman gain.

Initial conditions

- $B^j_{0|0} = (I_\ell - F_j)^{-1}\mu_j$
- $\mathrm{vec}(P^j_{0|0}) = (I_{\ell^2} - F_j \otimes F_j)^{-1}[G \otimes G]\,\mathrm{vec}\, Q_j$
- $Pr(s_0 = k) = \pi_k$ (steady-state probability)

Recall that $\mathcal{F}_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\}$ is the information set up to time $t-1$, $B^k_{t|t-1} = E(B_t|\mathcal{F}_{t-1}, s_{t-1} = k)$ is an inference on B_t based on $\mathcal{F}_{t-1}$ given $s_{t-1} = k$; $B^{(k,j)}_{t|t-1} = E(B_t|\mathcal{F}_{t-1}, s_t = j, s_{t-1} = k)$ is an inference on B_t based on $\mathcal{F}_{t-1}$, given $s_t = j$ and $s_{t-1} = k$; $P^k_{t-1|t-1}$ is the mean squared error matrix of $B^k_{t-1|t-1}$ conditional on $s_{t-1} = k$; $P^{(k,j)}_{t|t-1}$ is the mean squared error matrix of $B^{(k,j)}_{t|t-1}$ conditional on $s_t = j$ and $s_{t-1} = k$; $\eta^{(k,j)}_{t|t-1}$ is the conditional forecast error of y_t based on information up to time $t-1$, given

Table 1 Estimation results of simulated data from a 2-dimensional MS BEKK(1,1) model with $A_1 = A_2 = 0$ and $p_{11} = p_{22} = 0.5$ using Proposal 1.

Param	DGP	$n = 500$	$n = 1000$	$n = 2000$
$\tilde{C}_1$	$\begin{pmatrix} 0.20 & 0 \\ -0.56 & 0.20 \end{pmatrix}$	$\begin{pmatrix} 0.2541 & 0 \\ -0.1644 & 0.1027 \end{pmatrix}$	$\begin{pmatrix} 0.2727 & 0 \\ -0.1946 & 0.1344 \end{pmatrix}$	$\begin{pmatrix} 0.2895 & 0 \\ -0.2512 & 0.1447 \end{pmatrix}$
		[0.4110]	*[0.3783]*	*[0.3262]*
$\tilde{C}_2$	$\begin{pmatrix} 0.40 & 0 \\ -0.40 & 0.30 \end{pmatrix}$	$\begin{pmatrix} 0.0918 & 0 \\ -0.2520 & 0.0109 \end{pmatrix}$	$\begin{pmatrix} 0.1151 & 0 \\ -0.2944 & 0.0007 \end{pmatrix}$	$\begin{pmatrix} 0.1822 & 0 \\ -0.3518 & 0.0442 \end{pmatrix}$
		[0.4621]	*[0.4275]*	*[0.4102]*
B_1	$\begin{pmatrix} -0.37 & 0.07 \\ 0.26 & 0.22 \end{pmatrix}$	$\begin{pmatrix} 0.0168 & -0.0559 \\ 0.1470 & 0.0671 \end{pmatrix}$	$\begin{pmatrix} -0.0317 & -0.0225 \\ 0.1607 & 0.1048 \end{pmatrix}$	$\begin{pmatrix} -0.0902 & -0.0199 \\ 0.1867 & 0.1700 \end{pmatrix}$
		[0.4490]	*[0.3823]*	*[0.3070]*
B_2	$\begin{pmatrix} 0.50 & -0.05 \\ 0.38 & -0.12 \end{pmatrix}$	$\begin{pmatrix} 0.5892 & 0.4279 \\ 0.2729 & -0.1928 \end{pmatrix}$	$\begin{pmatrix} 0.5240 & 0.5165 \\ 0.2863 & -0.3285 \end{pmatrix}$	$\begin{pmatrix} 0.5648 & 0.5288 \\ 0.3043 & -0.4106 \end{pmatrix}$
		[0.5031]	*[0.6114]*	*[0.6553]*
p_{11}	0.50	0.6386	0.6360	0.6249
		[0.1386]	*[0.1360]*	*[0.1249]*
p_{22}	0.50	0.3316	0.3374	0.3576
		[0.1684]	*[0.1626]*	*[0.1424]*

Notes: The table reports the average estimation results of simulated data from a 2-dimensional regime switching BEKK(1,1) model based on 500 Monte Carlo simulations. The data generating process with true parameters is listed in the second column. The parameter estimates based on different sample sizes n=500, 1000 and 2000 are listed in the third, fourth and fifth column, respectively. RMSE is reported in square brackets below each estimated parameter matrix.

Table 2 Estimation results of simulated data from a 2-dimensional MS BEKK(1,1) model with $A_1 = A_2 = 0$ and $p_{11} = p_{22} = 0.5$ using Proposal 2.

Param	DGP	$n = 500$	$n = 1000$	$n = 2000$
$\tilde{C}_1$	$\begin{pmatrix} 0.20 & 0 \\ -0.56 & 0.20 \end{pmatrix}$	$\begin{pmatrix} 0.1624 & 0 \\ -0.4382 & 0.1313 \end{pmatrix}$	$\begin{pmatrix} 0.1637 & 0 \\ -0.4738 & 0.1355 \end{pmatrix}$	$\begin{pmatrix} 0.1696 & 0 \\ -0.4930 & 0.1444 \end{pmatrix}$
		[0.1448]	*[0.1136]*	*[0.0922]*
$\tilde{C}_2$	$\begin{pmatrix} 0.40 & 0 \\ -0.40 & 0.30 \end{pmatrix}$	$\begin{pmatrix} 0.4534 & 0 \\ -0.4068 & 0.3239 \end{pmatrix}$	$\begin{pmatrix} 0.4320 & 0 \\ -0.4190 & 0.3209 \end{pmatrix}$	$\begin{pmatrix} 0.4206 & 0 \\ -0.4184 & 0.3198 \end{pmatrix}$
		[0.0589]	*[0.0427]*	*[0.0340]*
B_1	$\begin{pmatrix} -0.37 & 0.07 \\ 0.26 & 0.22 \end{pmatrix}$	$\begin{pmatrix} -0.3480 & 0.0755 \\ 0.2184 & 0.2099 \end{pmatrix}$	$\begin{pmatrix} -0.3599 & 0.0742 \\ 0.2294 & 0.2141 \end{pmatrix}$	$\begin{pmatrix} -0.3628 & 0.0735 \\ 0.2349 & 0.2142 \end{pmatrix}$
		[0.0484]	*[0.0330]*	*[0.0270]*
B_2	$\begin{pmatrix} 0.50 & -0.05 \\ 0.38 & -0.12 \end{pmatrix}$	$\begin{pmatrix} 0.4865 & -0.0511 \\ 0.3712 & -0.1149 \end{pmatrix}$	$\begin{pmatrix} 0.4999 & -0.0511 \\ 0.3810 & -0.1194 \end{pmatrix}$	$\begin{pmatrix} 0.5004 & -0.0508 \\ 0.3817 & -0.1196 \end{pmatrix}$
		[0.0169]	*[0.0016]*	*[0.0020]*
p_{11}	0.50	0.5052	0.5022	0.5006
		[0.0052]	*[0.0022]*	*[0.0001]*
p_{22}	0.50	0.4955	0.4984	0.4981
		[0.0045]	*[0.0016]*	*[0.0019]*

Notes: The table reports the average estimation results of simulated data from a 2-dimensional regime switching BEKK(1,1) model based on 500 Monte Carlo simulations. The data generating process with true parameters is listed in the second column. The parameter estimates based on different sample sizes n=500, 1000 and 2000 are listed in the third, fourth and fifth column, respectively. RMSE is reported in square brackets below each estimated parameter matrix.

$s_t = j$ and $s_{t-1} = k$; and $f_{t|t-1}^{(k,j)}$ is the conditional variance of forecast error $\eta_{t|t-1}^{(k,j)}$.

Each iteration of the Kalman Filter produces a d-fold increase in the number of cases to consider. It is necessary to introduce some approximations to make the filter operable. The key is to collapse the $(d \times d)$ posteriors $B_{t|t}^{(k,j)}$ and $P_{t|t}^{(k,j)}$ into d posteriors $B_{t|t}^{j}$ and $P_{t|t}^{j}$. Hence, we mimic the approximation proposed by Kim (1994) (see also Kim and Nelson,

Table 3 Estimation results of simulated data from a bivariate MS(2) BEKK(1,1) model using Proposal 2.

Param	DGP	$n = 500$	$n = 1000$	$n = 2000$
$\tilde{C}_1$	$\begin{pmatrix} 0.20 & 0 \\ -0.56 & 0.20 \end{pmatrix}$	$\begin{pmatrix} 0.0644 & 0 \\ -0.1668 & 0.0782 \end{pmatrix}$	$\begin{pmatrix} 0.0803 & 0 \\ -0.2068 & 0.1271 \end{pmatrix}$	$\begin{pmatrix} 0.1114 & 0 \\ -0.2927 & 0.1481 \end{pmatrix}$
		[0.4334]	*[0.3800]*	*[0.2863]*
$\tilde{C}_2$	$\begin{pmatrix} 0.40 & 0 \\ -0.40 & 0.90 \end{pmatrix}$	$\begin{pmatrix} 0.3809 & 0 \\ -0.3546 & 0.8040 \end{pmatrix}$	$\begin{pmatrix} 0.4197 & 0 \\ -0.4009 & 0.8239 \end{pmatrix}$	$\begin{pmatrix} 0.4037 & 0 \\ -0.4030 & 0.8426 \end{pmatrix}$
		[0.1079]	*[0.0786]*	*[0.0576]*
A_1	$\begin{pmatrix} 0.38 & 0.06 \\ -0.9 & 0.35 \end{pmatrix}$	$\begin{pmatrix} 0.3111 & 0.0451 \\ -0.8185 & 0.3152 \end{pmatrix}$	$\begin{pmatrix} 0.3444 & 0.0530 \\ -0.8602 & 0.3431 \end{pmatrix}$	$\begin{pmatrix} 0.3686 & 0.0528 \\ -0.8872 & 0.3560 \end{pmatrix}$
		[0.1132]	*[0.0543]*	*[0.0195]*
A_2	$\begin{pmatrix} -0.11 & -0.04 \\ -0.06 & 0.01 \end{pmatrix}$	$\begin{pmatrix} -0.0992 & -0.0426 \\ -0.0611 & 0.0155 \end{pmatrix}$	$\begin{pmatrix} -0.1058 & -0.0402 \\ -0.0606 & 0.0155 \end{pmatrix}$	$\begin{pmatrix} -0.1080 & -0.0407 \\ -0.0623 & 0.0128 \end{pmatrix}$
		[0.0124]	*[0.0069]*	*[0.0042]*
B_1	$\begin{pmatrix} -0.37 & 0.07 \\ 0.26 & 0.22 \end{pmatrix}$	$\begin{pmatrix} -0.3210 & 0.0516 \\ 0.2617 & 0.2057 \end{pmatrix}$	$\begin{pmatrix} -0.3419 & 0.0622 \\ 0.2555 & 0.2120 \end{pmatrix}$	$\begin{pmatrix} -0.3578 & 0.0657 \\ 0.2594 & 0.2171 \end{pmatrix}$
		[0.0543]	*[0.0306]*	*[0.0133]*
B_2	$\begin{pmatrix} 0.70 & -0.05 \\ 0.38 & -0.02 \end{pmatrix}$	$\begin{pmatrix} 0.6600 & -0.0467 \\ 0.3652 & -0.0198 \end{pmatrix}$	$\begin{pmatrix} 0.6890 & -0.0487 \\ 0.3724 & -0.0180 \end{pmatrix}$	$\begin{pmatrix} 0.6990 & -0.0511 \\ 0.3776 & -0.0198 \end{pmatrix}$
		[0.0428]	*[0.0136]*	*[0.0028]*
p_{11}	0.80	0.7816	0.8001	0.8015
		[0.0184]	*[0.0001]*	*[0.0015]*
p_{22}	0.95	0.8877	0.8121	0.9323
		[0.0623]	*[0.1379]*	*[0.0177]*

Notes: The table reports the average estimation results of simulated data from a bivariate regime switching MS(2) BEKK(1,1) model over the 500 Monte Carlo simulations. The data generating process with true parameters is listed in the second column. The parameter estimates based on different sample sizes n=500, 1000 and 2000 are listed in the third, fourth and fifth column, respectively. RMSE is reported in square brackets below each estimated parameter matrix.

1999) applied to this state space representation. Let $B^j_{t|t}$ be the expectation based not only on $\mathcal{F}_t$ but also conditional on the random variable s_t taking on the value j, that is,

$$B^j_{t|t} = E[B_t \mid \mathcal{F}_t,\ s_t = j].$$

Then the approximation results to be

$$\begin{aligned} B^j_{t|t} &= \frac{\sum_{k=1}^d B^{(k,j)}_{t|t}\ Pr(s_{t-1} = k,\ s_t = j \mid \mathcal{F}_t)}{Pr(s_t = j|\mathcal{F}_t)} \\ &= \sum_{k=1}^d \frac{Pr(s_{t-1} = k,\ s_t = j \mid \mathcal{F}_t)}{Pr(s_t = j \mid \mathcal{F}_t)} B^{(k,j)}_{t|t} \\ &= \sum_{k=1}^d Pr(s_{t-1} = k \mid s_t = j,\ \mathcal{F}_t)\, B^{(k,j)}_{t|t} \\ &= \sum_{k=1}^d p^j_{k,t-1|t,t}\ B^{(k,j)}_{t|t} \end{aligned} \tag{32}$$

where

$$p^j_{k,t-1|t,t} := Pr(s_{t-1} = k|s_t = j, \mathcal{F}_t).$$

To justify the use of this approximation, note that the Kalman filter would give the conditional expectation if, conditional on $\mathcal{F}_{t-1}$ and on $s_t = j$, $s_{t-1} = k$, the distribution of B_t is normal. However, the distribution of B_t, conditional on $\mathcal{F}_{t-1}$ and on $s_t = j$, $s_{t-1} = k$, is a mixture of normal for $t > 2$. Hence, Kim (1994) proposes an approximation in which the exponential Gaussian mixture is collapsed down to d Gaussians at each step. This is a natural proxy for such a process.
Note that

$$p^j_{k,t-1|t,t} = \frac{p_{kj}\ p_{k,t-1|t-1}}{p_{j,t|t-1}}$$

where $p_{j,t|t-1} = Pr(s_t = j|\mathcal{F}_{t-1})$ is the prediction probability, $p_{k,t-1|t-1} = Pr(s_{t-1} = k|\mathcal{F}_{t-1})$ is the filtered probability and $p_{kj} = Pr(s_t = j|s_{t-1} = k)$ is the transition probability (assumed to be time invariant).

4 Numerical simulations

Initially we evaluate the filtering and estimation performance of the two filters proposed in Section 3 through Monte Carlo simulations. Let us consider a 2-dimensional MS BEKK(1,1) model

$$X_t = \Sigma_t^{1/2}(\mathcal{I}_{t-1}, s_t)\epsilon_t$$
$$\Sigma_t(\mathcal{I}_{t-1}, s_t) = \tilde{C}_{s_t}'\tilde{C}_{s_t} + A_{s_t}X_{t-1}X_{t-1}'A_{s_t}'$$
$$+ B_{s_t}\Sigma_{t-1}(\mathcal{I}_{t-2}, s_{t-1})B_{s_t}'$$

where $\epsilon_t \sim IID(0, I_2)$ and $s_t = 1, 2$. The parameter matrices are

$$A_{s_t} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}_{s_t}, \qquad B_{s_t} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}_{s_t},$$

where a_{12}, a_{21}, b_{12} and b_{22} indicates bidirectional volatility transmission between the series.

Engle and Kroner (1995) showed that the BEKK representation $\Sigma_t = C'C + \sum_{i=1}^{q} A_i X_{t-i} X_{t-i}' A_i' + \sum_{i=1}^{p} B_i \Sigma_{t-i} B_i'$ is identified if $\tilde{C}$ is lower triangular with positive diagonal elements. We follow this standard BEKK practice regime-wise to ensure the identification of the model parameters and the positive definiteness of $\tilde{C}_{s_t}'\tilde{C}_{s_t}$ via Cholesky decomposition.

First, we set $A_1 = A_2 = 0$ in the data generating process for a simplified 2-dimensional MS BEKK model. The rest true values of the parameters $\tilde{C}_1, \tilde{C}_2, B_1, B_2, p_{11}$ and p_{22} are listed in column 2 of Tables 1 and 2, for the first and second proposals, respectively. Three datasets, with different sample size, $n = 500$, 1000 and 2000 have been simulated from the above model. We estimate the model using the two proposed filters from Section 3, with 500 iterations considered for each sample size. The estimation results using the filter described in Proposal 1 and 2 are given in Tables 1 and 2, respectively.

The performance of the proposed filters is very satisfying, especially for proposal 2. Here we report the Root Mean Squared Error (RMSE) between the estimated and true parameter matrices. The RMSE is calculated using the Frobenius norm of their difference, divided by the square root of the matrix dimension to ensure proper normalization (Fan et al., 2013). In fact, for proposal 2, the RMSEs are very small and close to zero. The only exception is in the estimation of $\tilde{C}_1$ in which discrepancies are a little bit larger. For proposal 1, the RMSE of the matrices appear to be reasonably large. Moreover, the estimation results are more accurate as sample size increases for all the parameters, and consequently RMSE decrease as sample size increases.

Next, we consider a full 2-dimensional MS BEKK model. The true parameter values for $\tilde{C}_1, \tilde{C}_2, A_1, A_2, B_1$ and B_2 are provided in column 2 of Table 3 for the second proposal. These parameters are broadly consistent with our estimates from the real-data application in Section 4.2. The state variables are generated by a first order Markov process with the transition probability matrix

$$\boldsymbol{P} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} 0.80 & 0.20 \\ 0.05 & 0.95 \end{pmatrix}$$

which implies low persistence in regime 1 and high persistence in regime 2. Again, we simulate three datasets with sample size $n = 500$, 1000 and 2000, conducting 500 iterations for each sample size. The estimation results, obtained using the filter described in Proposal 2, are presented in Table 3. Despite the model's complexity, the proposed filter performs very well. It provides clear distinguish between the two sates and most of the estimations show small RMSE. Again, the estimation results are more accurate as sample size increases. Overall, the proposed filters exhibit appealing results and accurate estimation performance.

All of our simulations and estimations were made using Matlab(2024a) in a desktop (Intel Core i7CPU with 32 GB Ram). With this set up, the average computation time of one simulation with sample size $n = 1000$ is around 284s, for proposal 1 and 20s for proposal 2, respectively[2].

5 Real-data applications

We now turn to an empirical application of the multivariate regime switching BEKK(1,1) model to FTSE 100 and S&P 500 return data and compare it with two closely related alternatives in the literature, namely, the Markov switching constant conditional correlation (MS-CCC) model and the Markov switching dynamic conditional correlation model.

We consider daily log-returns on the FTSE100 and S&P 500, calculated as $r_{i,t} = 100(\ln(p_{i,t}) - \ln(p_{i-1,t}))$, $i = 1, 2$, for the period between 2 January 2010 to 31 December 2019. The samples of the S&P 500 and FTSE 100 consist of 2,514 observations each and the daily observations are presented in Figure 1[3]. From the plots it looks like the US stock market has a quicker speed of adjustment to its long-run equilibrium than that of UK stock market. The S&P 500 recovered strongly from the 2008 recession, while the performance of the FTSE 100 lagged behind that of the S&P 500 but exhibited similar increasing trends. Furthermore, from the movement of the returns in Figure 1, we can deduce that the downward movements of the S&P 500 and FTSE 100 tend to be associated with large returns.

Following Billio and Caporin (2005), we consider the MS-CCC and MS-DCC models as follows

$$X_t = \Sigma_t^{1/2}(\mathcal{I}_{t-1}, s_t)\epsilon_t$$

[2] Note that the average computation time of one simulation with sample size $n = 500$ is around 170s, for proposal 1 and 15s for proposal 2; with sample size $n = 2000$ is around 572s, for proposal 1 and 72s for proposal 2.

[3] The shaded part of this figure represents the out-of-sample period for our forecasting exercise later.

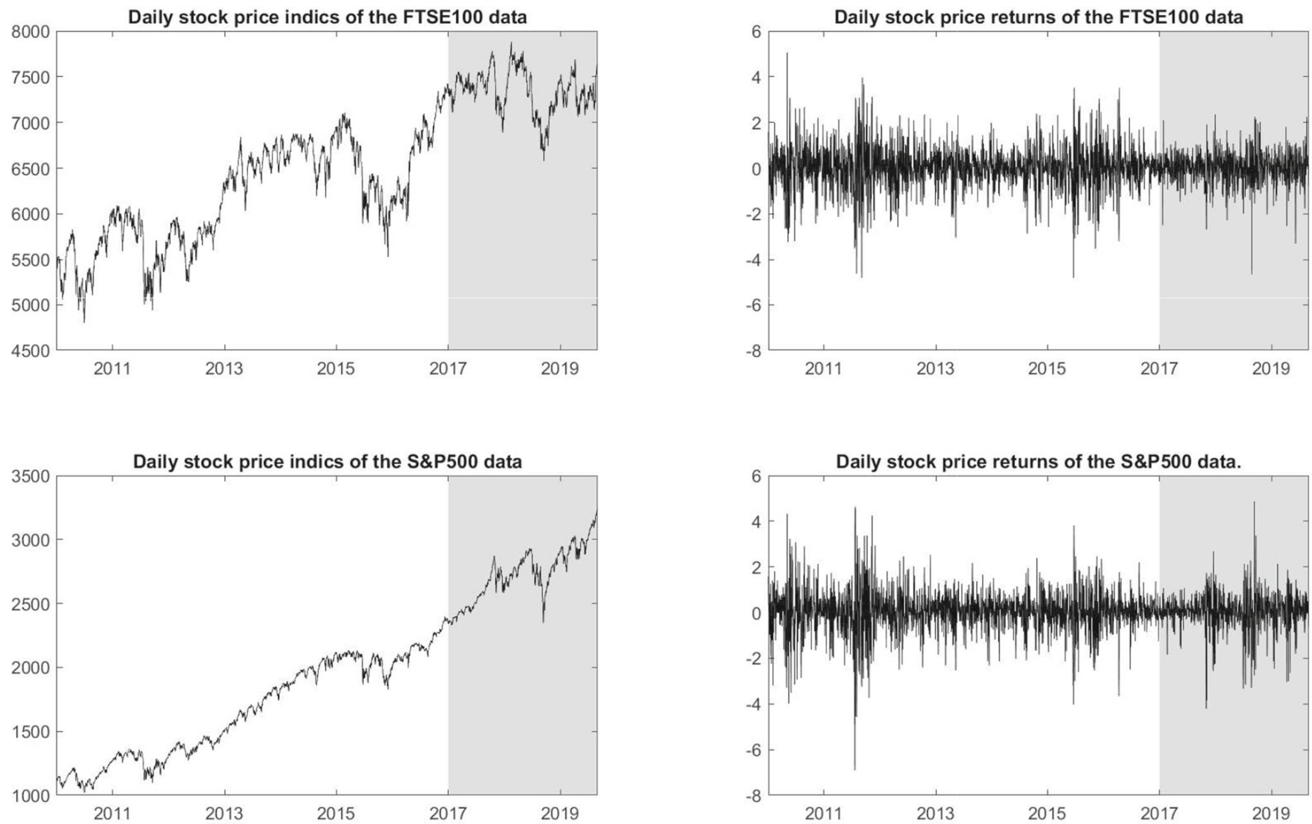

Fig. 1 Daily log-returns of FTSE100 and S&P 500.

Table 4 Estimation results of FTSE100 and S&P500 from MS-CCC and MS-DCC models.

Marginal Parameters	ω	α	β	
FTSE100	0.0456	0.1313	0.8233	
	[0.0093]	*[0.0151]*	*[0.0189]*	
S&P500	0.0451	0.1440	0.8071	
	[0.0077]	*[0.0220]*	*[0.0202]*	
Joint Parameters	a	b	ρ	$Pr(s_t\|s_{t-1})$
MS-CCC				
Regime 1			0.04859	0.8394
			[0.5007]	*[0.1002]*
Regime 2			0.8403	0.7873
			[0.1114]	*[0.1577]*
MS-DCC				
Regime 1	0.0524	0.3222	0.9969	0.2689
	[0.0744]	*[0.1201]*	*[0.0042]*	*[0.0614]*
Regime 2	0.0000	0.7668	0.3057	0.7447
	[0.0203]	*[0.1012]*	*[0.0647]*	*[0.0819]*

Notes: The table reports the estimation results of empirical data (FTSE100 and S&P500) from MS-CCC and MS-DCC models. The parameter estimates are based on equations (33), (34) and (35). Standard error is reported in square brackets below each estimated parameter.

Table 5 Estimation results of FTSE100 and S&P500 from a 2-dimensional MS BEKK(1,1) model.

Param	Proposal 1	Proposal 2
$\tilde{C}_1$	$\begin{pmatrix} 1.3869 & 0 \\ [0.0196] & \\ -0.6246 & 2.9480 \\ [0.0406] & [0.1710] \end{pmatrix}$	$\begin{pmatrix} 1.3424 & 0 \\ [0.0278] & \\ -0.6105 & 2.9821 \\ [0.0079] & [0.0235] \end{pmatrix}$
$\tilde{C}_2$	$\begin{pmatrix} 0.5172 & 0 \\ [0.0823] & \\ -0.1262 & 0.9882 \\ [0.1462] & [0.4290] \end{pmatrix}$	$\begin{pmatrix} 0.7141 & 0 \\ [0.6860] & \\ -0.7850 & 0.9836 \\ [0.5849] & [0.5494] \end{pmatrix}$
A_1	$\begin{pmatrix} 0.3954 & 0.0390 \\ [0.1180] & [0.1080] \\ -0.8990 & 0.3487 \\ [0.1014] & [0.1132] \end{pmatrix}$	$\begin{pmatrix} 0.3803 & 0.0450 \\ [0.0712] & [0.0251] \\ -0.8994 & 0.3428 \\ [0.0517] & [0.0274] \end{pmatrix}$
A_2	$\begin{pmatrix} -0.1003 & -0.0309 \\ [0.0084] & [0.0007] \\ -0.0505 & 0.0053 \\ [0.0811] & [0.0038] \end{pmatrix}$	$\begin{pmatrix} -0.0263 & 0.0494 \\ [0.0135] & [0.0124] \\ -0.0422 & 0.0006 \\ [0.0375] & [0.0227] \end{pmatrix}$
B_1	$\begin{pmatrix} -0.3983 & 0.0707 \\ [0.0157] & [0.0515] \\ 0.2607 & 0.1997 \\ [0.0274] & [0.0035] \end{pmatrix}$	$\begin{pmatrix} -0.3923 & 0.0793 \\ [0.0322] & [0.0078] \\ 0.2541 & 0.1953 \\ [0.0185] & [0.0131] \end{pmatrix}$
B_2	$\begin{pmatrix} 0.6988 & -0.0460 \\ [0.0618] & [0.0198] \\ 0.3056 & -0.0217 \\ [0.1427] & [0.0281] \end{pmatrix}$	$\begin{pmatrix} 0.5292 & 0.4573 \\ [0.0742] & [0.0230] \\ 0.2930 & 0.0420 \\ [0.0682] & [0.0324] \end{pmatrix}$
p_{11}	0.9503	0.9593
	[0.2467]	*[0.4283]*
p_{22}	0.9796	0.9801
	[0.1324]	*[0.0631]*

Notes: The table reports the estimation results of empirical data (FTSE100 and S&P 500) from a bivariate regime switching MS(2) BEKK(1,1) model. The standard errors are reports in square brackets.

$$\Sigma_t^{1/2}(\mathcal{I}_{t-1}, s_t) = D_t R_t(\mathcal{I}_{t-1}, s_t) D_t$$

where $D_t = \text{diag}(\sigma_{1t}, \sigma_{2t}, \cdots, \sigma_{nt})$ is a diagonal matrix of standard deviations and $R_t(\mathcal{I}_{t-1}, s_t)$ is a correlation matrix. The marginal variance equation for each log returns is

$$\sigma_t^2 = \omega + \alpha x_{t-1}^2 + \beta \sigma_{t-1}^2. \tag{33}$$

The standard MS-CCC model has the correlation matrix

$$R_t = \begin{pmatrix} 1, & \rho_{12,s_t}, & \cdots, & \rho_{1n,s_t} \\ \rho_{12,s_t}, & 1, & \cdots, & \rho_{2n,s_t} \\ \vdots, & \vdots, & \ddots, & \vdots \\ \rho_{1n,s_t}, & \rho_{2n,s_t} & \cdots, & 1 \end{pmatrix} \tag{34}$$

while in the MS-DCC model, the correlation matrix follows a time dependent relation as follows

$$\begin{aligned} R_t &= Q_t^{*-\frac{1}{2}} Q_t Q_t^{*-\frac{1}{2}} \\ Q_t &= (1 - a_{s_t} - b_{s_t})\overline{Q}_{s_t} + a_{s_t}\eta_{t-1}\eta_{t-1}' + b_{s_t} Q_{t-1} \\ Q_t^* &= \text{diag}(\sqrt{q_{11,t}}, \sqrt{q_{22,t}}, \cdots, \sqrt{q_{nn,t}}) \\ \overline{Q}_{s_t} &= \begin{pmatrix} 1, & \rho_{12,s_t}, & \cdots, & \rho_{1n,s_t} \\ \rho_{12,s_t}, & 1, & \cdots, & \rho_{2n,s_t} \\ \vdots, & \vdots, & \ddots, & \vdots \\ \rho_{1n,s_t}, & \rho_{2n,s_t} & \cdots, & 1 \end{pmatrix} \end{aligned} \tag{35}$$

where $\overline{Q}$ represents the unconditional correlations, Q_t^* guarantees that R_t is a correlation matrix ($q_{ii,t}, i = 1, 2, \cdots, n$ are the elements of the diagonal of Q_t), $\eta_t = D_t^{-1} X_t$.

5.1 In-sample estimation results

To assess both in-sample and out-of-sample performances of the three competing models, we split the whole sample into two parts. We use the first 70% of the sample (1762 observations, from January 2010 to December 2016) for in-sample analysis and the remaining 30% (752 observations, from Jan 2017 to Dec 2019) as our forecasting sample.

Table 4 summarizes our in-sample estimation results for the MS-CCC and MS-DCC models, where the standard error of each estimate is provided in square brackets beneath. The models are estimated by a two-step Quasi Maximum likelihood approach via the modified Markov switching filter, as demonstrated by Billio and Caporin (2005). All univariate GARCH parameters are significant at the 5% level. For the MS-CCC model, the transition probabilities indicate that the regimes are persistent, with one representing a high correlation state (Regime 2) and the other a (nearly) zero correlation state (Regime 1). In contrast, the MS-DCC model shows that, while dynamic correlation is accounted for, the regimes classification weakens. There is only one persistent regime (Regime 2) with a lower correlation coefficient.

The bivariate regime switching BEKK model was structured to allow for possible bidirectional volatility spillovers across different markets and estimated using the methods based on proposals 1 and 2. The parameter estimates for the MS(2) BEKK(1,1) model are reported in Table 5. The overall conditional volatilities, conditional covariance and conditional correlation for both time series are plotted in Figure 2. For comparison purpose, we also provided the conditional volatilities, conditional covariance and conditional correlation for each state in Figure 3 and 4 where it can be observed significant time-varying volatility and correlation for both FTSE 100 and S&P 500.

Overall, the estimation results using proposal 1 and 2 are very close to each other. The parameters a_{11}, b_{11}, a_{22} and b_{22} are all significant at the 5% level for state 1, indicating the volatilities of these two indices are influenced by their own past shocks and past volatilities. This result can be further

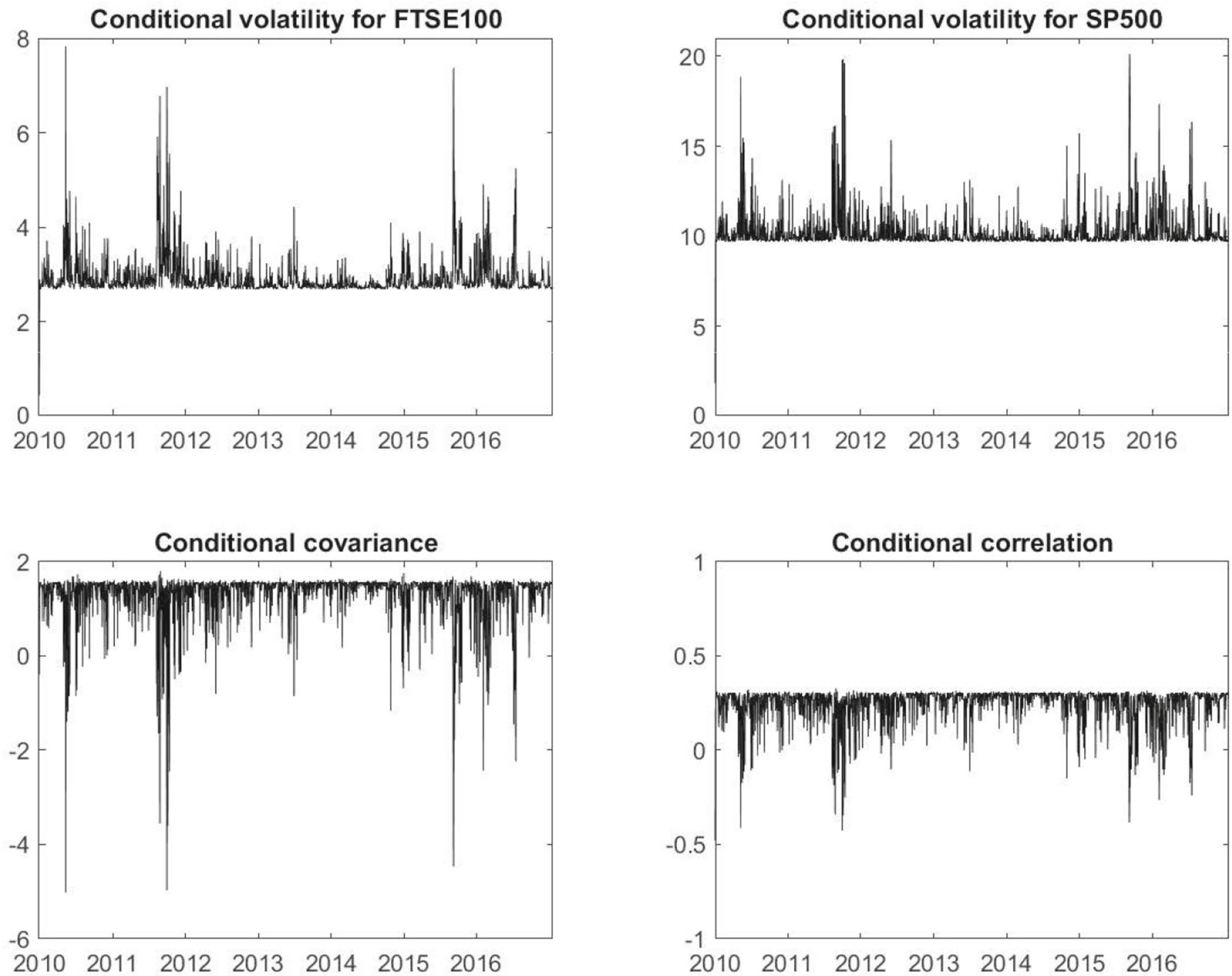

Fig. 2 Conditional volatility and correlation

confirmed by the conditional volatilities in Figure 2. However, some of these parameters are not significant for state 2 which represented more stable period. This is not surprising, as we can see from Figure 4, the volatilities of S&P 500 varies only between 1.5 to 2.5. Moreover, the parameters a_{12}, a_{21}, b_{12} and b_{21} are not significant for state 2, which indicates that there is no unidirectional effect between FTSE 100 and S&P 500. On the other side, the parameters a_{21} and b_{21} are strongly significant, hence we can say the volatility of the S&P 500 has a great effect on the fluctuation in the FTSE 100 return during economic turmoil.

Furthermore, to investigate the direction of volatility spillovers between the FTSE 100 and S&P 500 across both regimes, we conduct Wald test on the parameters estimated using proposal 2[4]. This procedure follows standard practice in multivariate GARCH literature (see, e.g., Engle and Kroner, 1995; Hafner and Herwartz, 2006; Cappiello et al., 2006). The results are presented in Table 6. For state 2, we find no evidence of volatility spillovers in either direction. In contrast, for state 1, we observe a statistically significant volatility spillover from the S&P 500 to the FTSE 100, while only weak evidence supports spillovers in the reverse direction.

The filtered probabilities of state 1 for three models are presented in Figure 5. As expected, the MS-CCC and MS-DCC models are unable to provide clear regime classification. For the MS-BEKK model, we present the filtered probabilities of state 1 using proposal 2[5]. Here state 1 represented the high volatility periods while state 2 is related to more stable period. In fact, from bottom plot in Figure 5 we recognize several economic, financial and political shocks: the 2010 European sovereign debt crisis and the stock Market

[4] Same conclusion can be drawn for proposal 1 using Wald test.

[5] Both proposals yield nearly identical filtered probabilities.

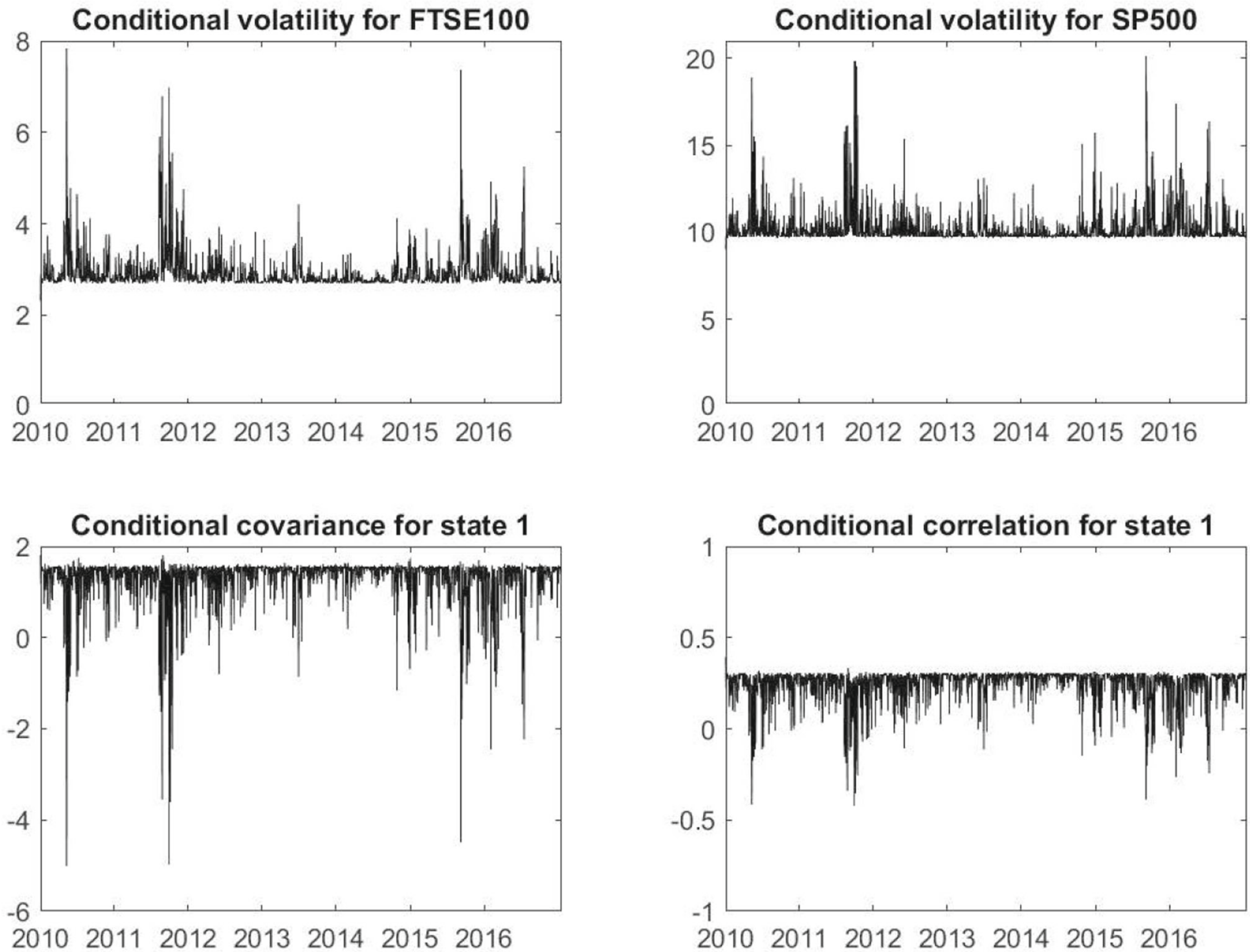

Fig. 3 Conditional volatility and correlation for state 1

Fall in August 2011; then the US Operation Twist announced in September 2011 lasting until September 2012, a short debt-ceiling crisis in January 2013 and finally the 2015-16 US stock market sell. The overall conditional volatilities, conditional covariance and conditional correlation for both time series are plotted in Figure 2.

5.2 Out-of-sample forecasting results

To complete our analysis, we present a forecasting exercise for the conditional volatility of our returns series. Concentrating on the risk measure of VaR, we analyze the quality of daily VaR forecasts based on the conditional density forecasts derived from the competing models discussed above. We compare the VaR prediction performance of the MS BEKK model to its various special cases, such as the MS-CCC and MS-DCC models using a rolling window scheme. The length of the rolling estimation window is set to be 752 observations, such that 752 observations (from January 3 2017, until December 31 2019) are left for out-of-sample forecast evaluation.

The backtest is performed for the one-day-ahead daily 5%, 1% and 0.5% VaR forecasts of the equally weighted portfolio. The results are presented in Table 7, which provides the unconditional coverage (UC) test (Kupiec, 1995) and conditional coverage (CC) test (Christoffersen, 1998) statistics and the corresponding p-values. The asymptotic distribution is a decent approximation because we use a large sample size of 752 VaR forecasts. For all models, the number of violations occurs at the expected frequencies across all levels of VaR forecasts. For example, the MS-CCC model yields 30 violations out of 752 in the 95% VaR forecasts, resulting in a coverage rate of 0.0399 (approximately 5% exceedance), which aligns with the expected proportion 0.05.

For both the 1% and 0.5% VaR, all the models pass the backtest for all conventional confidence levels, though for the 0.5% VaR, the MS-CCC and MS-DCC models give marginal p-values of 0.0569 for the UC test. Meanwhile, for the 5%

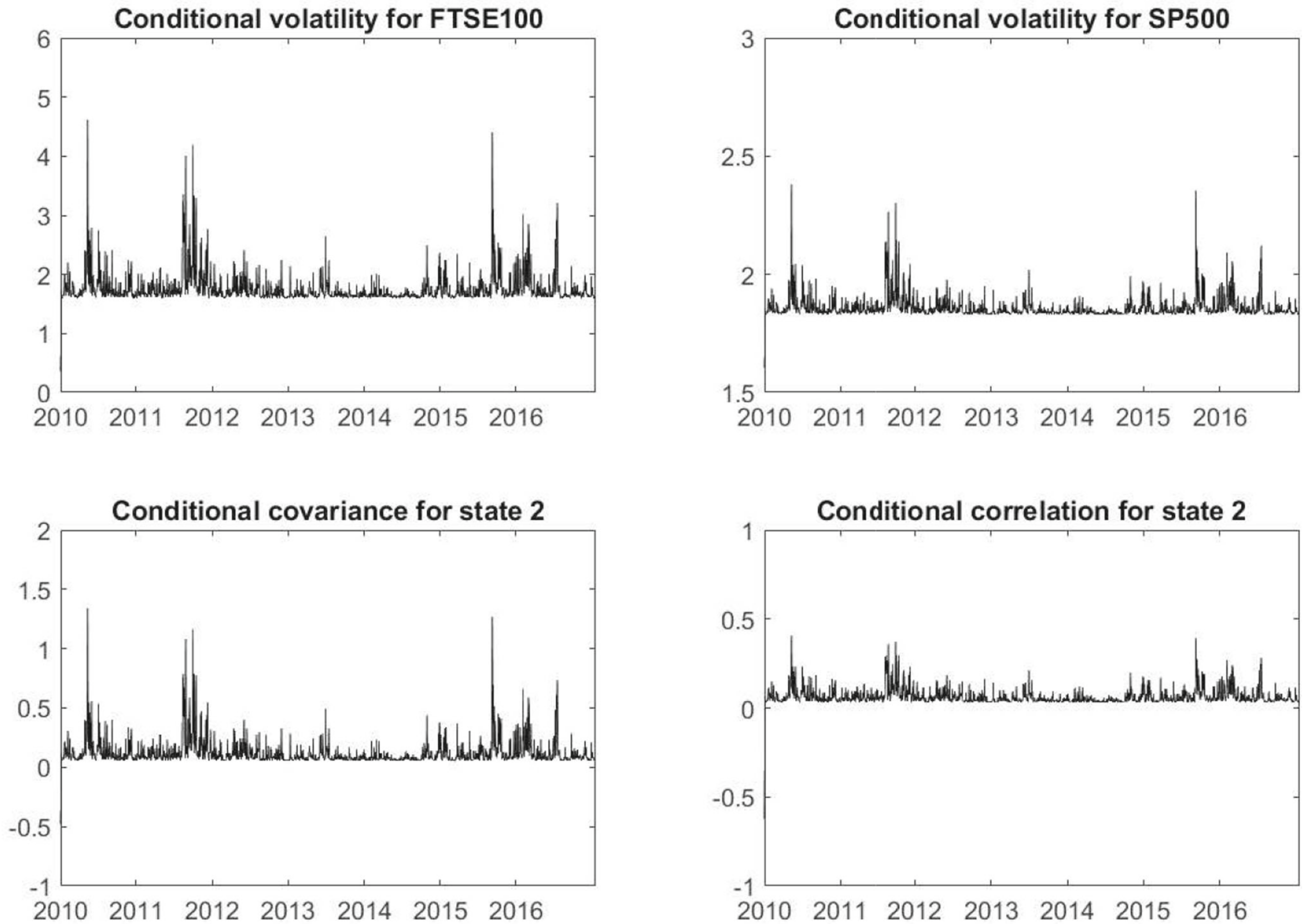

Fig. 4 Conditional volatility and correlation for state 2

Table 6 Wald test results of volatility spillover effect.

Null hypothesis	Wald test statistic	p-value
state 1		
H_0: $a_{12} = b_{12} = 0$	0.0188	0.9906
H_0: $a_{21} = b_{21} = 0$	10.9670	0.0042
state 2		
H_0: $a_{12} = b_{12} = 0$	0.0480	0.9763
H_0: $a_{21} = b_{21} = 0$	0.0125	0.9938

Notes: The number 1 and 2 in the subscript represent the returns of the FTSE 100 and S&P 500. State 1 represents the economic turmoil and state 2 represents the economic recovery.

VaR, the MS-BEKK model outperforms the MS-CCC and MS-DCC model in the CC test, as this is a critical measure, though it gives a marginal p-value for the UC test.

Then we show in Figure 6 the one-day-ahead daily 5%, 1% and 0.5% VaR estimates against the realized returns for the equally weighted portfolio. We observe that typically the VaR estimates based on the MS-BEKK models can characterize the risk well during periods of high market volatility, particularly when encountering substantial return fluctuations. In contrast, the MS-CCC and MS-DCC models provide more conservative VaR estimates that closely track the actual returns during relatively stable market conditions.

Another object of interest is the out-of-sample forecasts of regime probabilities, i.e. the one-step-ahead forecasts of which regime will occur next. Figure 5 presents the out-of-sample forecasts of the probabilities governing whether state 1 will occur next in the shaded area for the MS-CCC, MS-DCC and MS-BEKK models. Comparing the three plots, we observe that the forecasted state probabilities under the MS-CCC and MS-DCC models are either around 0.6 or consistently 0.3. This indicates that the probability of switching from one state to another is very low. In contrast, the results in the MS-BEKK models exhibit significant regime identification.

Remarkably, the forecasts of changes in market regimes inferred from the MS-BEKK model are closely linked to historical economic events in the US and UK stock markets. In the bottom plot of Figure 5, we can identify the time frame from 2018 to 2019 due to a) the uncertainties

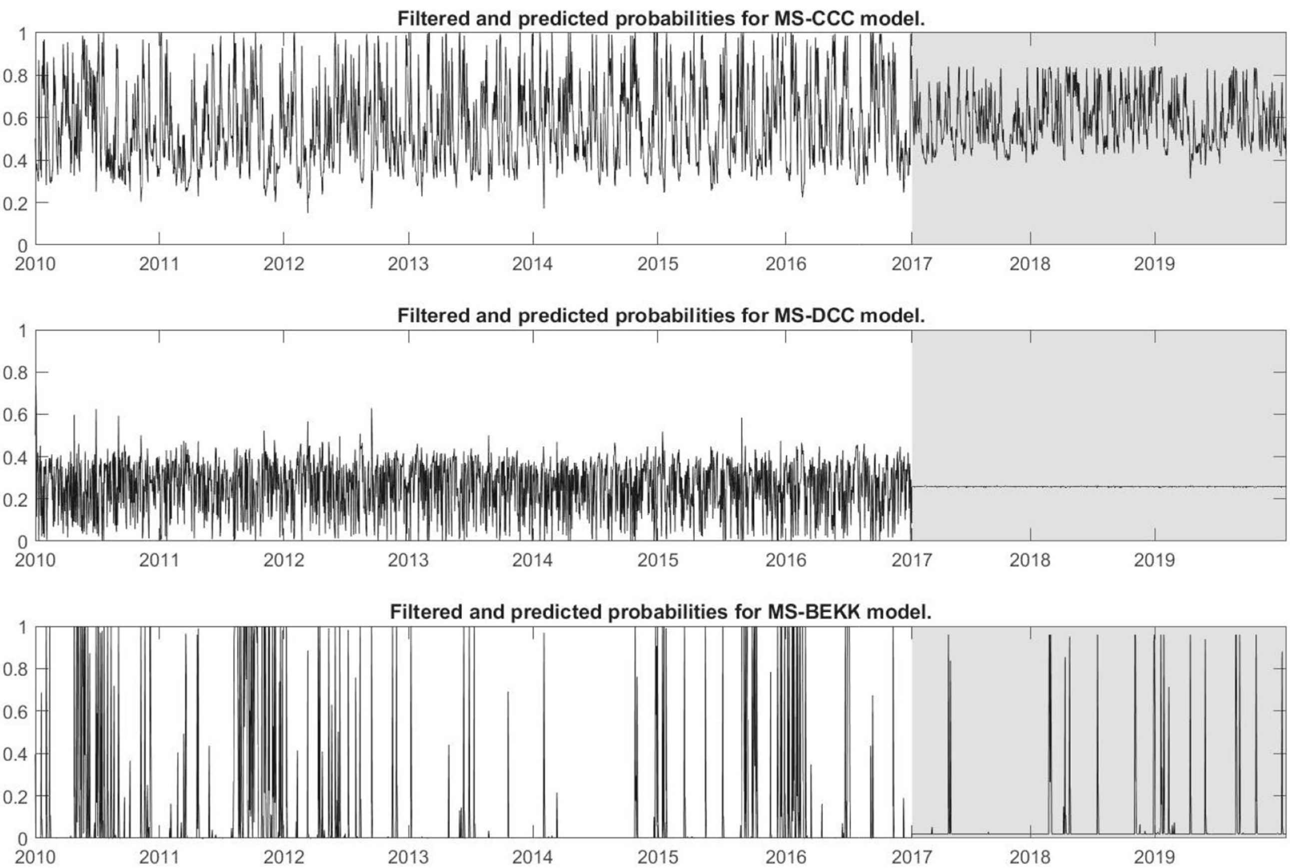

Fig. 5 Filtered and predicted probabilities of state 1 using MS-CCC, MS-DCC and MS-BEKK models. *Notes:* This figure shows the filtered probabilities and the predicted probabilities (shaded area) of state 1 for the MS-CCC model (top plot), the MS-DCC model (middle plot) and the MS-BEKK model (bottom plot).

Table 7 Results of out-of-sample VaR forecasting performance.

	CC	*p*-value	UC	*p*-value	violation	coverage-rate
5% VaR						
MS-CCC	4.6142	**0.0317**	1.7322	0.1881	30	0.0399
MS-DCC	7.1809	**0.0074**	1.2937	0.2554	31	0.0412
MS-BEKK	0.0115	0.9416	4.2042	**0.0403**	26	0.0346
1% VaR						
MS-CCC	0.2183	0.6403	0.2768	0.5988	9	0.0120
MS-DCC	0.2699	0.6034	0.7486	0.3869	10	0.0133
MS-BEKK	0.2183	0.6403	0.2768	0.5988	9	0.0120
0.5% VaR						
MS-CCC	0.1723	0.6781	3.6244	0.0569	8	0.0106
MS-DCC	0.1723	0.6781	3.6244	0.0569	8	0.0106
MS-BEKK	0.0966	0.7559	1.1348	0.2867	6	0.0080

Notes: Backtest results of one-day-ahead VaR forecasts of the equally weighted portfolio returns are obtained using MS-CCC, MS-DCC and MS-BEKK models. The table provides the unconditional coverage (UC) and conditional coverage (CC) test statistics and the corresponding *p*-values. Bold numbers highlights *p*-values that are smaller than 5%. A violation occurs when the portfolio loss exceeds the VaR forecast.

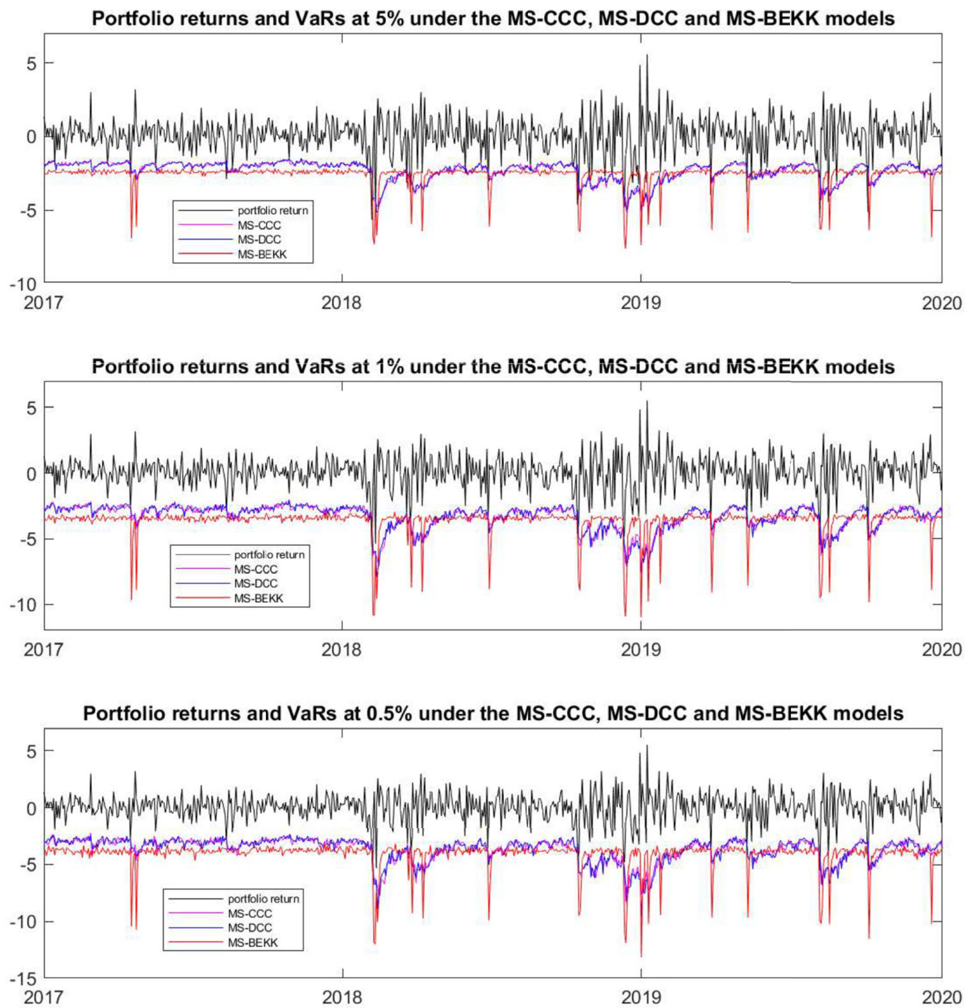

Fig. 6 Portfolio returns and VaRs under the MS-CCC, MS-DCC and MS-BEKK models. *Notes:* This figure shows the one-day-ahead daily 95%, 99% and 99.5% VaR forecasts of the equally weighted portfolios.

caused by the beginning of the trade war between China and the USA, whose consequences could negatively impact the returns to the S&P500, and even the correlation between the stock indices and b) Brexit negotiation imbroglios, i.e. the disagreement between the British government and the European bloc regarding the costs of the exit increased the risk of a unilateral exit from the British side (Hard Brexit), which kept the uncertainty in the markets.

In summary, our empirical analysis demonstrates that the proposed MS-BEKK model exhibits significant advantages in both in-sample estimation and out-of-sample forecasting performances. The model shows particular strength in precise identification of regime switching dynamics and accurate forecasting of volatility and correlation patterns in stock index data. These advantages stem from its flexible matrix specification, which allows for comprehensive modelling of volatility spillovers and regime-specific asymmetric covariance dynamics.

Finally, the MS-BEKK framework is especially valuable during periods of market stress (e.g., the 2010 European sovereign debt crisis and the debt-ceiling crisis in 2013), where its ability to capture contagion effects through regime-dependent spillovers leads to superior performance. Furthermore, the model produces more accurate Value-at-Risk (VaR) forecasts than alternative specifications, as evidenced by backtesting results. This enhanced forecasting capability has direct practical implications for portfolio optimization and risk management applications.

6 Conclusion

We consider the extension of the classical multivariate BEKK model from Engle and Kroner (1995) to the class of Markov switching (MS) BEKK models by allowing the state-dependent dynamics. We then derive a MS VARMA representation and different state space descriptions of the proposed model. These are used to present two alternative filtering methods for parameter and state estimation based on extended Kalman filter. We compare the derived filtering algorithms using simulation experiments with Gaussian innovations and we conclude that the filter algorithm in proposal 2 outperforms the algorithm in proposal 1 in terms of computation time and accuracy. Finally, we discuss some empirical applications to real-life data which make evident the effectiveness and relevance of the proposed filtering and nonlinear estimation.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11222-026-10897-7.

Acknowledgements This work is financially supported by research grants of the University of Modena and Reggio Emilia, Italy, and Keele University, UK. We are grateful to the Editor of the journal, Professor Ajay Jasra, and the two anonymous referees for their very useful suggestions and remarks which greatly improved the final version of the paper.

Author Contributions Maddalena Cavicchioli and Jie Cheng contributed equally to this work.

Funding Open access funding provided by Università degli Studi di Modena e Reggio Emilia within the CRUI-CARE Agreement.

Data availability Data are publicly available.

Declarations

Competing interests The authors declare no competing interests.

References

Abramson, A., Cohen, I.: On the stationarity of Markov-switching GARCH processes. Economet. Theor. **23**(3), 485–500 (2007)

Alexander, C., Lazar, E.: Modelling regime-specific stock price volatility. Oxford Bull. Econ. Stat. **71**, 761–797 (2009)

Ang, A., Timmermann, A.: Regime changes and financial markets. Annu. Rev. Financ. Econ. **4**, 313–337 (2012)

Aradia, D., Bluteau, K., Rüede, M.: Regime changes in Bitcoin GARCH volatility dynamics. Financ. Res. Lett. **29**, 266–271 (2019)

Ardia, D., Bluteau, K., Boudt, K., Catania, L.: Forecasting risk with Markov switching GARCH models: A large-scale performance study. Intern. J. Forecasting **34**(4), 733–747 (2018)

Augustyniak, M.: Maximum likelihood estimation of the Marko-switching GARCH model. Computational Statistics & Data Analysis **76**, 61–75 (2014)

Ausin, M.C., Galeano, P.: Bayesian estimation of the Gaussian mixture GARCH model. Computational Statistics & Data Analysis **51**(5), 2636–2652 (2007)

Bauwens, L., Hafner, C.M., Rombouts, J.V.K.: Multivariate mixed normal conditional heteroskedasticity. Computational Statistics & Data Analysis **51**, 3551–3566 (2007)

Bauwens, L., Preminger, A., Rombouts, J.V.K.: Theory and inference for a Markov switching GARCH model. Econometrics Journal **13**, 218–244 (2010)

Billio, M., Caporin, M.: Multivariate Markov switching dynamic conditional correlation GARCH representations for contagion analysis. Stat. Methods Appl. **14**, 145–161 (2005)

Billio, M., Casarin, R., Osuntuyi, A.: Efficient Gibbs sampling for Markov switching GARCH models. Computational Statistics & Data Analysis **100**, 37–57 (2016)

Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics **31**, 307–327 (1986)

Bollerslev, T.: Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH approach. Rev. Econ. Stud. **72**, 498–505 (1990)

Bollerslev, T.: Glossary to ARCH (GARCH). in: Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle (eds. T. Bollerslev, J.R. Russell and M. Watson), Oxford Univ. Press, Oxford, UK (2008)

Boussama, F., Fuchs, F., Stelzer, R.: Stationarity and geometric ergodicity of BEKK multivariate GARCH models. Stochastic Processes and their Applications **121**(10), 2331–2360 (2011)

Bu, R., Cheng, J., Hadri, K.: Specification analysis in regime-switching continuous-time diffusion models for market volatility. Studies in Nonlinear Dynamic & Econometrics **21**(1), 65–80 (2017)

Cappiello, L., Engle, R.F., Sheppard, K.: Asymmetric dynamics in the correlations of global equity and bond returns. J. Financ. Economet. **4**(4), 537–572 (2006)

Cavicchioli, M.: Fourth moment structure of Markov switching multivariate GARCH models. J. Financ. Economet. **19**(4), 565–582 (2021)

Cavicchioli, M.: Markov switching GARCH models: Higher order moments, kurtosis measures, and volatility evaluation in recessions and pandemic. Journal of Business & Economic Statistics **40**(4), 1772–1783 (2022)

Cavicchioli, M.: Spectral analysis of Markov switching GARCH models with statistical inference. Scand. J. Stat. **50**(1), 102–119 (2023)

Chkili, W.: Modeling Bitcoin price volatility: long memory vs Markov switching. Eurasian Econ. Rev. **11**, 433–448 (2021)

Engle, R.F.: Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. Econometrica **50**, 987–1007 (1982)

Engle, R.F.: Dynamic conditional correlation: A new simple class of multivariate GARCH models. Journal of Business & Economic Statistics **20**(3), 339–350 (2002)

Engle, R.F., Kroner, K.F.: Multivariate simultaneous generalized ARCH. Economet. Theor. **11**(1), 122–150 (1995)

Fan, J., Liao, Y., Mincheva, M.: Large covariance estimation by thresholding principal orthogonal complements. J. Roy. Stat. Soc. B **75**(4), 603–680 (2013)

Foroni, B., Merlo, L., Petrella, L.: Expectile hidden Markov regression models for analyzing cryptocurrency returns. Stat. Comput. **34**(2), 66 (2024)

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

Francq, C., Zakoïan, J.M.: The L^2-structures of standard and switching-regime GARCH models. Stochastic Processes and their Applications **115**, 1557–1582 (2005)

Francq, C., Zakoïan, J.M.: Deriving the autocovariances of powers of Markov-switching GARCH models with applications to statistical inference. Computational Statistics & Data Analysis **52**(6), 3027–3046 (2008)

Francq, C., Zakoïan, J.M.: QML estimation of a class of multivariate asymmetric GARCH models. Economet. Theor. **28**, 179–206 (2012)
Frühwirth-Schnatter, S.: Finite Mixture and Markov Switching Models. Springer Verlag, New York (2006)
Ghorbel, A., Jeribi, A.: Volatility spillovers and contagion between energy sector and financial assets during COVID-19 crisis period. Eurasian Econ. Rev. **11**, 449–467 (2021)
Gourieroux, C., Monfort, A.: Time Series and Dynamic Models. Cambridge Univ. Press, Cambridge (1997)
Guidolin, M.: Markov switching models in empirical finance, pp. 1–86. Time-series Methods and Applications. Emerald Group Publishing Limited, In Missing Data Methods (2011)
Guidolin, M., Hyde, S.: Simple VARs cannot approximate Markov switching asset allocation decisions: An out-of-sample assessment. Computational Statistics & Data Analysis **56**(11), 3546–3566 (2012)
Haas, M., Liu, J.-C.: A multivariate regime-switching GARCH model with an application to global stock market and real estate equity returns. Studies in Nonlinear Dynamics & Econometrics **22**(3), 20160019 (2018)
Hafner, C.M.: Fourth moment structure of multivariate GARCH models. J. Financ. Economet. **1**(1), 26–54 (2003)
Hafner, C.M., Herwartz, H.: Volatility impulse responses for multivariate GARCH models: An exchange rate illustration. J. Int. Money Financ. **25**(5), 719–740 (2006)
Hafner, C.M., Preminger, A.: On asymptotic theory for multivariate GARCH models. J. Multivar. Anal. **100**, 2044–2054 (2009)
Hamilton, J.D.: A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica **57**, 357–384 (1989)
Hamilton, J.D.: Analysis of time series subject to changes in regime. Journal of Econometrics **45**, 39–70 (1990)
Hamilton, J.D.: Time Series Analysis. Princeton University Press, Princeton, N.J. (1994)
Hamilton, J.D.: *State-Space Models*. Handbook of Econometrics vol. IV, Elsevier Science, 3039–3080 (1994b)
Hamilton, J.D.: Chapter 3–Macroeconomic Regimes and Regime Shifts. in: Handbook of Macroeconomics (Elsevier Publ.) vol.2, 163–201 (2016)
Hu, L., Shin, Y.: Optimal test for Markov switching GARCH models. *Studies in Nonlinear Dynamics & Econometrics* 12 (3) (2008)
Jang, Y., Sundararajan, R.R., Barreto-Sousa, W.: A multivariate heavy-tailed integer-valued GARCH process with EM algorithm-based inference. Stat. Comput. **34**(1), 56 (2024)
Jha, K.K., Baur, D.G.: Regime-dependent good and bad volatility of Bitcoin. Journal of Risk and Financial Management **13**, 1–16 (2020)
Kim, C.-J.: Dynamic linear models with Markov switching. Journal of Econometrics **60**, 1–22 (1994)
Kim, C.-J., Nelson, C.R.: State-Space Models with Regime Switching. The MIT Press, Cambridge, MA, Classical and Gibbs-Sampling Approaches with Applications (1999)
Krolzig, H.M.: Markov-switching Vector Autoregressions: Modelling. Springer-Verlag, Berlin-Heidelberg-New York, Statistical Inference and Application to Business Cycle Analysis (1997)
Lee, H.-T.: The effects of asymmetries and regime switching on optimal futures hedging. Applied Financial Economics Letters **4**(2), 133–136 (2008)
Lee, H.-T., Yoder, J.K.: A bivariate Markov regime switching GARCH approach to estimate time-varying minimum variance hedge ratios. Appl. Econ. **39**(10), 1253–1265 (2007)
Ling, S., McAleer, M.: Asymptotic theory for a vector ARMA-GARCH model. Economet. Theor. **19**, 280–310 (2003)
Liu, J.-C.: Stationarity of a Markov-switching GARCH model. J. Financial Econometrics **4**(4), 573–593 (2006)
Livingston, G.C., Jr., Nur, D.: Bayesian inference of multivariate GARCH-BEKK models. Stat. Pap. **64**(5), 1749–1774 (2023)
Magnus, J.R., Neudecker, H.: The commutation matrix: some properties and applications. Ann. Stat. **7**, 381–394 (1979)
Magnus, J.R., Neudecker, H.: Symmetry, $0-1$ Matrices and Jacobians: A Review. Economet. Theor. **2**, 157–190 (1986)
Nielsen Rotta, P., Valls Pereira, P.L.: Analysis of contagion from the dynamic conditional correlation model with Markov regime switching. Appl. Econ. **48**(25), 2367–2382 (2016)
Pasricha, G.K.: Kalman filter and its economic applications. MPRA paper No.22734 (2006)
Sajjad, R., Coakley, J., Nankervis, J.C.: Markov switching GARCH modelling of value-at-risk. *Studies in Nonlinear Dynamics & Econometrics* 12 (3) (2008)
Su, Y.K., Wu, C.C.: A new range-based regime-switching dynamic conditional correlation model for minimum-variance hedging. Journal of Mathematical Finance **4**(3), 207 (2014)
Tan, C.Y., Koh, Y.B., Ng, K.H.: Dynamic volatility modelling of Bitcoin using time-varying transition probability Markov-switching GARCH model. The North American Journal of Economics and Finance **56**, 101377 (2021)
Xuan, H., Maestrini, L., Chen, F., Grazian, C.: Stochastic variational inference for GARCH models. Stat. Comput. **34**(1), 45 (2024)
Zhiping, Y., Shenghong, L.: Hedge ratio on Markov regime-switching diagonal BEKK-GARCH model. Financ. Res. Lett. **24**, 49–55 (2018)
Zhu, Y., Cheng, J.: Using hidden Markov model to detect macroeconomic risk level. Review of Integrative Business and Economics Research **2**(1), 238 (2013)