Journal of Financial Econometrics, 2021, Vol. 19, No. 4, 565–582
doi: 10.1093/jjfinec/nbz020
Advance Access Publication Date: 6 July 2019
Article

OXFORD

Fourth Moment Structure of Markov Switching Multivariate GARCH Models

Maddalena Cavicchioli

University of Modena and Reggio Emilia

Address correspondence to Maddalena Cavicchioli, Department of Economics "Marco Biagi", University of Modena and Reggio Emilia, V.le Berengario 51, Modena, 41121, Italy, or e-mail: maddalena.cavicchioli@unimore.it.

Received April 3, 2018; revised April 30, 2019; editorial decision May 1, 2019; accepted May 4, 2019

Abstract

We derive sufficient conditions for the existence of second and fourth moments of Markov switching multivariate generalized autoregressive conditional heteroscedastic processes in the general vector specification. We provide matrix expressions in closed form for such moments, which are obtained by using a Markov switching vector autoregressive moving-average representation of the initial process. These expressions are shown to be readily programmable in addition of greatly reducing the computational cost. As theoretical applications of the results, we derive the spectral density matrix of the squares and cross products, propose a new definition of multivariate kurtosis measure to recognize heavy-tailed features in financial real data, and provide a matrix expression in closed form of the impulse-response function for the volatility. An empirical example illustrates the results.

Key words: Markov switching models, conditional heteroscedasticity, multivariate Markov switching GARCH models, Markov switching VARMA representations, fourth moments, spectral density, multivariate kurtosis, volatility

JEL classification: C01, C05, C32

Since the seminal works of Engle (1982) and Bollerslev (1986), generalized autoregressive conditional heteroscedastic (GARCH) models have been frequently used in modeling the volatility of financial time series. These dynamic models are shown to be more capable than standard time series models, such as vector autoregressive moving-average (VARMA) processes, to capture the empirical features in several financial data, such as stock prices or indices, and exchange rates. Estimation, consistency, and asymptotic theory of quasi-maximum likelihood (QML) estimators of the parameters of multivariate GARCH models are provided by Comte and Lieberman (2003), Hafner and Preminger (2009), and Francq and Zakoïan (2012). Stationarity conditions for GARCH models in various specifications can be found in Francq and Zakoïan (2011, §10).

A feature of the GARCH model is that the conditional variance changes over time as a fixed function of the past. However, it has been shown empirically that many financial time series typically exhibit structural changes in the dynamics of the conditional variance, which are not accounted for by standard GARCH models. One popular approach in modeling changes in regime is to consider Markov switching (MS) parameters. Several authors have proposed MS GARCH models to model the volatility in financial time series. For information concerning stationarity, consistency of maximum likelihood (ML) estimates geometric ergodicity, L^2-structure, filtering, duality, and statistical inference of univariate MS GARCH models, see Francq, Roussignol, and Zakoïan (2001), Francq and Zakoïan (2005), Liu (2006), and Bauwens, Preminger, and Rombouts (2010).

A special case of MS GARCH models are the so-called mixture GARCH models, where the regime variable is identically and independently distributed (i.i.d.) across different dates. In the univariate GARCH case, mixture models have been introduced by Haas, Mittnik, and Paolella (2004), while Bauwens, Hafner, and Rombouts (2007) extend such models to the multivariate case. Both papers derive the fourth moment structure, so there is some link with the results obtained in this paper. An advantage of mixture models is that in high dimensions, simple models with few parameters can be mixed to obtain more flexibility than specifying a complex one-component model. On the other hand, introducing changes in regime substantially increases the model flexibility. This makes MS models more preferable than mixture ones in several empirical economic applications. In the mixing case, it is possible that covariance stationarity, and hence finite second moments, holds although for some components it does not hold, see Bauwens, Hafner, and Rombouts (2007). Something like this is also possible in the MS case, see Bauwens, Preminger, and Rombouts (2010).

Our goal is to study the second and fourth moment structures of multivariate MS GARCH models in the general vector specification [including, the BEKK specification of Engle and Kroner (1995)]. The approach to do this is based on an MS VARMA representation of the initial process. We generalize the matrix formula obtained by Hafner (2003) for the second moments of standard multivariate GARCH models to the MS context. Then we derive a matrix expression in closed form for the fourth moment structure of multivariate MS GARCH models. This formula is new, nonrecursive, and without the use of an infinite summation as in Hafner (2003), so it is easily tractable and directly computable. Estimating GARCH-type models often requires to calculate their unconditional moments, which are needed for variance targeting. To our knowledge, there are no explicit matrix formulas in closed form, which allow to do this analytically for the class of multivariate MS GARCH processes. Our matrix expressions for the second and fourth unconditional moments of multivariate MS GARCH models are shown to be readily programmable. So they greatly reduce the computational cost of estimating unconditional moments for such models.

As a first application, we propose a new definition of multivariate kurtosis measure for multivariate MS GARCH models. As remarked by Hafner (2003), this is relevant for various purposes. For example, having estimated the model based on QML, the question arises as to which innovation distribution one should use in simulation studies such that the empirical kurtosis is approximated. In finance, this is particularly important for option pricing, where the degree of excess kurtosis explains the shape of the so-called smile. A second example is the estimation by generalized method of moments, where the efficiency may often be improved using the kurtosis formula. Further theoretical applications of our results include the derivation of the spectral density matrix of the squares and cross products, and a matrix expression in closed form of the impulse-response function for the volatility of multivariate MS GARCH models.

The article is structured as follows. Section 1 is devoted to introduce the model and state the main results. In Section 2, we propose the kurtosis measure and derive the spectral density matrix of the squares and cross products and the impulse-response function for the volatility of a multivariate MS GARCH model. Section 3 applies the results to a bivariate MS GARCH. Section 4 concludes. Some proofs are given in the Appendix. Additional proofs and examples are provided in the Supplementary Material.

1 Assumptions and Main Results

Let us consider the general M-state MS m-dimensional GARCH(p, q) model [in short, MS(M) GARCH(p, q)]:

$$\mathbf{x}_t = \mathbf{H}_t^{1/2}\boldsymbol{\eta}_t, \tag{1}$$

$$\mathbf{h}_t = \mathbf{c}(s_t) + \sum_{i=1}^{q}\mathbf{A}_i(s_t)\mathbf{y}_{t-i} + \sum_{j=1}^{p}\mathbf{B}_j(s_t)\mathbf{h}_{t-j}, \tag{2}$$

where $\mathbf{x}_t$ is a random vector with values in $\mathbb{R}^m$, $\mathbf{y}_t = \mathrm{vech}(\mathbf{x}_t\mathbf{x}_t') \in \mathbb{R}^K$ with $K = m(m+1)/2$, $\mathbf{H}_t = \mathbf{H}(\mathcal{F}_{t-1}, \Delta_t) \in \mathbb{R}^{m\times m}$, and $\mathbf{h}_t = \mathrm{vech}(\mathbf{H}_t) \in \mathbb{R}^K$. Here, $\mathcal{F}_t$ denotes the information set available at time t, that is, $\mathcal{F}_t = \{\mathbf{x}_t, \mathbf{x}_{t-1}, \ldots\}$ and $\Delta_t = \{s_t, s_{t-1}, \ldots\}$.

Assumption 1. *The process (s_t) is an irreducible, aperiodic, and ergodic Markov chain with values in the set $\Xi = \{1, 2, \ldots, M\}$, stationary transition probabilities $p_{ij} = Pr(s_t = j|s_{t-1} = i)$* for *$i, j = 1, \ldots, M$, and unconditional (or steady state) probabilities $\pi_i = Pr(s_t = i)$ for $i \in \Xi$.*

Let $\mathbf{P} = (p_{ij})$ denote the transition probability matrix of the chain. For the model parameters, we have $\mathbf{A}_i(s_t)$, $\mathbf{B}_j(s_t) \in \mathbb{R}^{K\times K}$ and $\mathbf{c}(s_t) \in \mathbb{R}^K$, where $\mathbb{R}^{m\times n}$ denotes the class of real $m \times n$ matrices and $\mathbb{R}^n$ the class of n-dimensional real vectors, that is, $\mathbb{R}^n = \mathbb{R}^{n\times 1}$.

Assumption 2. *The innovation $(\boldsymbol{\eta}_t)$ is i.i.d. with mean zero and identity covariance matrix. Furthermore, $(\boldsymbol{\eta}_t)$ is independent of (s_t).*

To simplify computations, the conditional mean of $(\mathbf{x}_t)$ is assumed to be zero in Equation (1). But the results of the article can be easily generalized to a nonzero mean case. Thus, we have $E(\mathbf{x}_t) = E(\mathbf{x}_t|\mathcal{F}_{t-1}) = E(\mathbf{x}_t|\mathcal{F}_{t-1}, \Delta_t) = \mathbf{0}$ and $\mathrm{var}(\mathbf{x}_t|\mathcal{F}_{t-1}, \Delta_t) = \mathbf{H}_t$. The square root of the matrix $\mathbf{H}_t$ in Equation (1) can be defined as $\mathbf{H}_t^{1/2} = \mathbf{Q}_t\Lambda_t^{1/2}\mathbf{Q}_t'$ with the matrix $\mathbf{Q}_t$ containing the eigenvectors and the diagonal matrix Λ_t containing the eigenvalues of $\mathbf{H}_t$ on its diagonal. To define the square root of $\mathbf{H}_t$, one can also use a Cholesky factorization of $\mathbf{H}_t$. A sufficient condition for $\mathbf{H}_t$ to be positive definite is that each of the parameter matrices is symmetric positive definite. The symmetry of the parameter matrices, however, is a rather restrictive assumption. There are as yet no general results on necessary and sufficient conditions for positivity of $\mathbf{H}_t$. This is why in practice one often considers restrictions that ensure positivity.

For example, the MS BEKK(p, q) model specifies $\mathbf{H}_t$ as

$$\mathbf{H}_t = \overline{\mathbf{C}}(s_t)\overline{\mathbf{C}}'(s_t) + \sum_{i=1}^{q}\overline{\mathbf{A}}_i(s_t)\mathbf{x}_{t-i}\mathbf{x}_{t-i}'\overline{\mathbf{A}}_i'(s_t) + \sum_{j=1}^{p}\overline{\mathbf{B}}_j(s_t)\mathbf{H}_{t-j}\overline{\mathbf{B}}_j'(s_t),$$

where $\overline{\mathbf{C}}(s_t)$ is a $m \times m$ lower triangular parameter matrix, $\overline{\mathbf{A}}_i(s_t)$ and $\overline{\mathbf{B}}_j(s_t)$ are $m \times m$ parameter matrices. See Engle and Kroner (1995) for the state-invariant case. Such a specification ensures positivity of $\mathbf{H}_t$. Moreover, it will typically involve fewer parameters than the vec equation in Equation (2). Positivity constraints on the conditional variances in the family of conditional correlation GARCH models can also be found in Nakatani and Teräsvirta (2008).

In this article, we derive results for the more general vec model in Equation (2), for which the following holds:

Assumption 3. *The matrix* $\mathbf{H}_t$ *is positive definite almost surely.*

Models (1) and (2) are the MS extension of the multivariate GARCH considered in Hafner (2003).

By rearranging terms, the model can be represented as an MS(M) VARMA(r, p), where $r = \max(p, q)$:

$$\mathbf{y}_t = \mathbf{c}(s_t) + \sum_{i=1}^{r} \mathbf{a}_i(s_t)\mathbf{y}_{t-i} + \boldsymbol{\varepsilon}_t + \sum_{j=1}^{p} \mathbf{b}_j(s_t)\boldsymbol{\varepsilon}_{t-j}, \tag{3}$$

where

$$\begin{aligned}
\boldsymbol{\varepsilon}_t &= \boldsymbol{\varepsilon}(\mathcal{F}_{t-1}, \Delta_t) = \mathbf{y}_t - \mathbf{h}_t = \boldsymbol{\sigma}_t \mathbf{u}_t \in \mathbb{R}^K \\
\boldsymbol{\sigma}_t &= \boldsymbol{\sigma}(\mathcal{F}_{t-1}, \Delta_t) = \mathbf{D}_m^+\left(\mathbf{H}_t^{1/2} \otimes \mathbf{H}_t^{1/2}\right) \in \mathbb{R}^{K\times m^2} \\
\mathbf{u}_t &= \boldsymbol{\eta}_t \otimes \boldsymbol{\eta}_t - \mathrm{vec}(\mathbf{I}_m) \in \mathbb{R}^{m^2}.
\end{aligned}$$

In order to give explicit results, we assume a specific distribution for the innovations $\boldsymbol{\eta}_t$. See Hafner (2003), Assumption 3.

Assumption 4. *The distribution of* $\boldsymbol{\eta}_t$ *belongs to the class of spherical distributions with finite fourth moments.*

Spherical distributions include the multinormal and multivariate Student's t-distributions. It is known that very often the latter distributions are used to accommodate fat tails of the innovation distribution.

If $\boldsymbol{\eta}_t$ satisfies Assumption 4 with mean zero, covariance identity matrix $\mathbf{I}_m$, and cokurtosis coefficient c, then the innovation process $(\mathbf{u}_t)$ is i.i.d. with mean zero and covariance matrix

$$\begin{aligned}
\Omega &= 2c\mathbf{D}_m\mathbf{D}_m^+ + (c-1)\mathrm{vec}(\mathbf{I}_m)\mathrm{vec}(\mathbf{I}_m)' \\
&= c[\mathbf{I}_{m^2} + \mathbf{K}_{mm}] + (c-1)\mathrm{vec}(\mathbf{I}_m)\mathrm{vec}(\mathbf{I}_m)'.
\end{aligned}$$

See Magnus and Neudecker (1986), Lemma 8, and Hafner (2003), Lemma 3. Here, $\mathbf{D}_m$ is the $m^2 \times K$ duplication matrix, $\mathbf{D}_m^+$ denotes the Moore–Penrose inverse of $\mathbf{D}_m$, that is, $\mathbf{D}_m^+ = \left(\mathbf{D}_m'\mathbf{D}_m\right)^{-1}\mathbf{D}_m' \in \mathbb{R}^{K\times m^2}$, and $\mathbf{K}_{mm}$ is the $m^2 \times m^2$ commutation matrix. For example, for a multinormal distribution $c = 1$, and for a multivariate Student's t-distribution with ν degrees of freedom, $c = (\nu-2)/(\nu-4)$ if $\nu > 4$.

In the MS VARMA representation (3), the $K \times K$ matrices $\mathbf{a}_i(s_t)$ are given by $\mathbf{a}_i(s_t) = \mathbf{A}_i(s_t) + \mathbf{B}_i(s_t)$, $i = 1, \ldots, r$, where we set $\mathbf{A}_{q+1}(s_t) = \cdots = \mathbf{A}_p(s_t) = \mathbf{0}$ if $p > q$ and

$\mathbf{B}_{p+1}(s_t) = \cdots = \mathbf{B}_q(s_t) = \mathbf{0}$ if $q > p$. Furthermore, $\mathbf{b}_j(s_t) = -\mathbf{B}_j(s_t)$ for $j = 1, \ldots, p$. Notice that $(\boldsymbol{\varepsilon}_t)$ is a martingale difference sequence such that $E(\boldsymbol{\varepsilon}_t) = E(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}) = \mathbf{0}$ and $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_\tau) = \mathbf{0}$ for $t \neq \tau$. However, $(\boldsymbol{\varepsilon}_t)$ is not independent over time since it does not have a constant variance in time, that is, it is not a homoscedastic process. Thus, $(\boldsymbol{\varepsilon}_t)$ is only weak white noise. This is the reason why Equation (3) allows only a study of covariance stationarity and not strict stationarity.

To investigate covariance stationarity of the process $(\mathbf{y}_t)$, we use a Markovian representation of Equation (3), proposed by Francq and Zakoïan (2001) (here, we set $n = K(p + r)$):

$$\mathbf{z}_t = \Phi_t \mathbf{z}_{t-1} + \boldsymbol{\omega}_t \qquad \boldsymbol{\omega}_t = \mathbf{c}_t + \Sigma_t, \tag{4}$$

where

$$\begin{aligned}
\mathbf{z}_t &= \begin{pmatrix} \mathbf{y}'_t & \mathbf{y}'_{t-1} & \cdots & \mathbf{y}'_{t-r+1} & \boldsymbol{\varepsilon}'_t & \boldsymbol{\varepsilon}'_{t-1} & \cdots & \boldsymbol{\varepsilon}'_{t-p+1} \end{pmatrix}' \in \mathbb{R}^n \\
\mathbf{c}_t &= \begin{pmatrix} \mathbf{c}'(s_t) & \mathbf{0}' & \cdots & \mathbf{0}' \end{pmatrix}' \in \mathbb{R}^n \\
\Sigma_t &= \Sigma(\mathcal{F}_{t-1}, \Delta_t) = \begin{pmatrix} \boldsymbol{\varepsilon}'_t & \mathbf{0}' & \cdots & \mathbf{0}' & \boldsymbol{\varepsilon}'_t & \mathbf{0}' & \cdots & \mathbf{0}' \end{pmatrix}' = \mathbf{g}\boldsymbol{\varepsilon}_t \in \mathbb{R}^n \\
\mathbf{g}' &= (\mathbf{I}_K \mathbf{0} \cdots \mathbf{0} \mathbf{I}_K \mathbf{0} \cdots \mathbf{0}) \in \mathbb{R}^{K \times n}
\end{aligned}$$

and

$$\Phi_t = \begin{pmatrix}
\mathbf{a}_1(s_t) & \mathbf{a}_2(s_t) & \cdots & \mathbf{a}_{r-1}(s_t) & \mathbf{a}_r(s_t) & \mathbf{b}_1(s_t) & \mathbf{b}_2(s_t) & \cdots & \mathbf{b}_{p-1}(s_t) & \mathbf{b}_p(s_t) \\
\mathbf{I}_K & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_K & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{I}_K & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_K & \cdots & \mathbf{0} & \mathbf{0} \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_K & \mathbf{0}
\end{pmatrix}$$

belongs to $\mathbb{R}^{n \times n}$. In the vectorial representation (4), it is implicitly assumed that $p \geq 1$ (hence, $r \geq 1$) without loss of generality because $\mathbf{b}_p(s_t)$ can be equal to $\mathbf{0}$ in Equation (3). Let $\Phi(i)$ and $\mathbf{c}(i)$ be the matrices obtained replacing s_t by i in Φ_t and $\mathbf{c}_t$, respectively. Then the following matrices can be defined:

$$\mathbf{P}(\Phi) = \begin{pmatrix}
p_{11}\Phi(1) & p_{21}\Phi(1) & \cdots & p_{M1}\Phi(1) \\
p_{12}\Phi(2) & p_{22}\Phi(2) & \cdots & p_{M2}\Phi(2) \\
\vdots & \vdots & & \vdots \\
p_{1M}\Phi(M) & p_{2M}\Phi(M) & \cdots & p_{MM}\Phi(M)
\end{pmatrix} \in \mathbb{R}^{(Mn) \times (Mn)}$$

and

$$\mathbf{c} = (\pi_1 \mathbf{c}'(1) \qquad \cdots \qquad \pi_M \mathbf{c}'(M))' \in \mathbb{R}^{Mn}.$$

Assumption 5. *All eigenvalues of the matrix* $\mathbf{P}(\Phi)$ *have modulus smaller than 1.*

Theorem 1. *Under Assumptions 1–3 and 5, the second moments of the multivariate* MS(M) GARCH(p,q) *process* $\mathbf{x}=(\mathbf{x}_t)$ *in* Equations (1) *and* (2) *are finite. In that case, the unconditional covariance matrix* $\Sigma_{\mathbf{x}}=\mathrm{var}(\mathbf{x}_t)$ *is given by*

$$\boldsymbol{\sigma}_{\mathbf{x}}=\mathrm{vech}(\Sigma_{\mathbf{x}})=(\mathbf{e}'\otimes\mathbf{f}')\mathbf{U},$$

where

$$\mathbf{U}=(\mathbf{I}_{Mn}-\mathbf{P}(\Phi))^{-1}\mathbf{c}.$$

Here, we set $\mathbf{e}=(1\cdots1)'\in\mathbb{R}^M$ *and* $\mathbf{f}'=(\mathbf{I}_K\quad \boldsymbol{0}\cdots\boldsymbol{0})\in\mathbb{R}^{K\times n}$.

The matrix expression for $\boldsymbol{\sigma}_{\mathbf{x}}$ in Theorem 1 generalizes formula (5) of Hafner (2003) obtained for multivariate GARCH models without regime switching.

The following result has been proved by Hafner (2003, theorem 1) for multivariate standard GARCH models. It also maintains its validity in the MS context (as shown in the Appendix), and serves to prove Theorem 3 below on the computation of the fourth moments for multivariate MS GARCH processes.

Theorem 2. *Set* $\Sigma_{\mathbf{y}}=E(\mathbf{y}_t\mathbf{y}_t')$, $\Sigma_{\mathbf{h}}=E(\mathbf{h}_t\mathbf{h}_t')$ *and* $\Sigma_{\boldsymbol{\varepsilon}}=E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')$. *Under Assumptions 1–5, we have*

$$\begin{aligned}\Sigma_{\boldsymbol{\varepsilon}}&=\Sigma_{\mathbf{y}}-\Sigma_{\mathbf{h}}\\ \mathrm{vec}(\Sigma_{\mathbf{y}})&=\mathbf{G}_K\,\mathrm{vec}(\Sigma_{\mathbf{h}})\\ \mathrm{vec}(\Sigma_{\boldsymbol{\varepsilon}})&=(\mathbf{G}_K-\mathbf{I}_{K^2})\,\mathrm{vec}(\Sigma_{\mathbf{h}})\end{aligned},$$

where

$$\mathbf{G}_K=c\{2(\mathbf{L}_m\otimes\mathbf{D}_m^+)(\mathbf{I}_m\otimes\mathbf{K}_{mm}\otimes\mathbf{I}_m)(\mathbf{D}_m\otimes\mathbf{D}_m)+\mathbf{I}_{K^2}\}$$

with $\mathbf{L}_m$ *denoting the elimination matrix and* $c=E[\eta_{1t}^4]/3$, *where* $\boldsymbol{\eta}_t=(\eta_{1t}\cdots\eta_{mt})'$.

Let $\mathbf{P}(\Phi\otimes\Phi)$ be defined replacing $\Phi(i)$ by $\Phi(i)\otimes\Phi(i)$ in the definition of $\mathbf{P}(\Phi)$. Then $\mathbf{P}(\Phi\otimes\Phi)$ is $(Mn^2)\times(Mn^2)$. Let $\mathbf{C}$ be defined replacing $\mathbf{c}(i)$ by $\mathbf{c}(i)\otimes\mathbf{c}(i)$ in the definition of $\mathbf{c}$. Then $\mathbf{C}$ is $(Mn^2)\times1$. Let $\mathbf{D}$ be defined replacing $\Phi(i)$ by $\mathbf{c}(i)\otimes\Phi(i)+\Phi(i)\otimes\mathbf{c}(i)$ in the definition of $\mathbf{P}(\Phi)$. Then $\mathbf{D}$ is $(Mn^2)\times(Mn)$.

Assumption 6. *All eigenvalues of the* $(Mn^2)\times(Mn^2)$ *matrix* $\mathbf{P}(\Phi\otimes\Phi)$ *have modulus smaller than 1, and the* $(Mn^2)\times(MK^2)$ *matrix*

$$\mathbf{Q}=(\mathbf{I}_M\otimes\tilde{\mathbf{f}})-[\mathbf{I}_{Mn^2}-\mathbf{P}(\Phi\otimes\Phi)]^{-1}(\mathbf{I}_M\otimes\tilde{\mathbf{g}})$$

has rank MK^2, *where* $\tilde{\mathbf{f}}=(\mathbf{f}\otimes\mathbf{f})\mathbf{G}_K$ *and* $\tilde{\mathbf{g}}=(\mathbf{g}\otimes\mathbf{g})(\mathbf{G}_K-\mathbf{I}_{K^2})$.

Notice that the first sentence of Assumption 6 implies Assumption 5, but the converse is not true in general. See, for example, Cavicchioli (2017).

Theorem 3. *Under Assumptions 1–4 and 6, the fourth moments of the multivariate* MS(M) GARCH(p,q) *process* $\mathbf{x}=(\mathbf{x}_t)$ *in* Equations (1) *and* (2) *are finite. In that case, the unconditional fourth moment of* $\mathbf{x}$ *is given by*

$$\mathrm{vech}(\Sigma_{\mathbf{y}})=\mathbf{D}_K^+\mathbf{G}_K(\mathbf{e}'\otimes\mathbf{I}_{K^2})\mathbf{V},$$

where $\mathbf{y} = (\mathbf{y}_t)$ *with* $\mathbf{y}_t = \mathrm{vech}\left(\mathbf{x}_t\mathbf{x}_t'\right)$, *and*

$$\mathbf{V} = \left(\mathbf{Q}'\mathbf{Q}\right)^{-1}\mathbf{Q}'\left[\mathbf{I}_{Mn^2} - \mathbf{P}(\Phi \otimes \Phi)\right]^{-1}(\mathbf{DU} + \mathbf{C}).$$

Note that $\mathrm{vec}(\Sigma_{\mathrm{y}}) = \left(\mathbf{D}_m^+ \otimes \mathbf{D}_m^+\right)E\left[(\mathbf{x}_t)^{\otimes 4}\right]$, where $(\mathbf{x}_t)^{\otimes 4}$ denotes the Kronecker product of four copies of $\mathbf{x}_t$.

Remark 1. In the case of multivariate mixed GARCH models, the fourth moment structure has been derived by Bauwens, Hafner, and Rombouts (2007, formula 18). Their formula relates with that given in Theorem 3: both expressions for vec (Σ_{y}) have the matrix $\mathbf{G}_K$ on the left side, followed by matrix products that look alike.

Remark 2. As remarked in the introduction, the formula for the fourth moments in Theorem 3 is nonrecursive and without the use of an infinite summation as in Hafner (2003) for standard vector GARCH models. The difference of approach with respect to the cited paper in obtaining the results can be explained as follows. Hafner (2003) uses a vector moving-average representation of infinite order, in short VMA(∞), for the VARMA model in Equation (3) (without shifts in regime). Our method of computation is based on the Markovian representation (4) for the MS VARMA model in Equation (3).

The proof of Theorem 3 (see the Appendix) and Theorem 2 from Francq and Zakoïan (2001) imply the following result:

Theorem 4. *Under Assumptions 1–4 and 6, for all* $t \in \mathbb{Z}$, *the series*

$$\mathbf{z}_t = \sum_{k=1}^{\infty} \Phi_t\Phi_{t-1}\cdots\Phi_{t-k+1}\boldsymbol{\omega}_{t-k} + \boldsymbol{\omega}_t$$

converges in L^2 *and the* MS(M) VARMA(r,p) *process* $(\mathbf{y}_t)$, *defined as the block of the first* K *components of* $(\mathbf{z}_t)$, *is the unique nonanticipative second-order stationary solution of* Equation (4).

Recall that a process $(\mathbf{y}_t)$ is called *nonanticipative* if, for all $t \in \mathbb{Z}$, $\mathbf{y}_t$ is measurable with respect to the σ–field generated by $\{s_\tau, \boldsymbol{\varepsilon}_\tau : \tau \leq t\}$.

In the sequel, we illustrate some theoretical implications of our results.

2 Spectral Density, Kurtosis, and Impulse-Response Functions for the Volatility

2.1 The Spectral Density Matrix of the Squares and Cross Products

It might be useful to look at the frequency domain properties of the squares and cross products of a multivariate MS GARCH process rather than at the time domain. For example, the model specification process might compare the empirical spectral density with the theoretical counterpart of a fitted model. The spectral density matrix of a covariance stationary stochastic process is defined as the Fourier transform of the autocovariance function, that is, $\mathbf{F}(\omega) = (2\pi)^{-1}\sum_{\tau=-\infty}^{\infty}\Gamma(\tau)\exp(-i\tau\omega)$. We compute the spectral density matrix of the vector of squares and cross products, $\mathbf{y}_t = \mathrm{vech}\left(\mathbf{x}_t\mathbf{x}_t'\right)$, of the multivariate MS GARCH(p, q) process $\mathbf{x} = (\mathbf{x}_t)$. Under Assumptions $1-4$, we derive a closed-form matrix expression for the autocovariance function $\Gamma_{\mathrm{y}}(\tau)$ of $\mathbf{y} = (\mathbf{y}_t)$. For all $\tau \geq 0$, let $\mathbf{W}(\tau)$ be the

$(Mn^2) \times 1$ vector whose i-th block is $\pi_i E(\mathbf{z}_t \otimes \mathbf{z}_{t-\tau} | s_t = i)$, for $i = 1, \ldots, M$. Then $\mathbf{W}(0) = \mathbf{W} = (\mathbf{I}_M \otimes \tilde{\mathbf{f}})\mathbf{V}$, where $\mathbf{V}$ is given in Theorem 3. Let $\mathbf{P}(\mathbf{c})$ be the $(Mn) \times M$ matrix defined replacing $\Phi(i)$ by $\mathbf{c}(i)$ in the definition of $\mathbf{P}(\Phi)$. For $\tau > 0$, the following matrix relation is proved in the Supplementary Material, Section 1:

$$\mathbf{W}(\tau) = [\mathbf{P}(\Phi) \otimes \mathbf{I}_n]\mathbf{W}(\tau - 1) + \left[\left(\mathbf{P}(\mathbf{c})(\mathbf{P}')^{\tau-1}\right) \otimes \mathbf{I}_n\right]\mathbf{U}, \tag{5}$$

where $\mathbf{P}$ is the transition probability matrix and $\mathbf{U}$ is given in Theorem 1.

Theorem 5. *Under Assumptions 1–5, the autocovariance function $\Gamma_{\mathbf{y}}(\tau)$ of $\mathbf{y} = (\mathbf{y}_t)$ is given by*

$$\mathrm{vec}\Gamma_{\mathbf{y}}(\tau) = (\mathbf{e}' \otimes \mathbf{f}' \otimes \mathbf{f}')\mathbf{W}(\tau) - \boldsymbol{\sigma}_{\mathbf{x}} \otimes \boldsymbol{\sigma}_{\mathbf{x}},$$

where

$$\mathbf{W}(\tau) = \left\{[\mathbf{P}(\Phi)]^{\tau} \otimes \mathbf{I}_n\right\}(\mathbf{I}_M \otimes \tilde{\mathbf{f}})\mathbf{V} + \mathbf{W}^*(\tau)$$

and

$$\mathbf{W}^*(\tau) = \sum_{i=1}^{\tau} \left\{[\mathbf{P}(\Phi)]^{i-1} \otimes \mathbf{I}_n\right\}\left\{\left[\mathbf{P}(\mathbf{c})(\mathbf{P}')^{\tau-i}\right] \otimes \mathbf{I}_n\right\}\mathbf{U}.$$

Now the spectral density function of $\mathbf{y}$ follows from Theorem 5 (here, we set $\mathbf{A}^i = \mathbf{I}_n$ if $i = 0$ for an $n \times n$ matrix $\mathbf{A}$).

2.2 Multivariate Kurtosis for MS GARCH Models

Having matrix expressions for the unconditional second and fourth moments of the multivariate MS GARCH process $\mathbf{x} = (\mathbf{x}_t)$, we can introduce a new kurtosis measure for such time series, which changes in regime, and derive a matrix formula in closed form for it. A deviation from the Gaussian distribution can be reflected by this measure, which depends on the characteristics of the volatility process. Let $\Sigma_{\mathbf{x}}$ be the covariance matrix of $\mathbf{x}$. Let us consider the process $\mathbf{x}^* = (\mathbf{x}_t^*)$ defined by

$$\mathbf{x}_t^* = \Sigma_{\mathbf{x}}^{-1/2}\mathbf{x}_t,$$

where $\Sigma_{\mathbf{x}}^{1/2}$ is any symmetric square root of $\Sigma_{\mathbf{x}}$. For the process $\mathbf{x}$, driven by the multivariate MS GARCH model in Equations (1) and (2), the *kurtosis* of $\mathbf{x}$ is defined to be the $K^2 \times 1$ vector, with $K = m(m+1)/2$, given by

$$\mathbf{k}(\mathbf{x}) := \mathrm{vec}\Sigma_{\mathbf{y}^*},$$

where $\mathbf{y}^* = (\mathbf{y}_t^*)$ and $\mathbf{y}_t^* = \mathrm{vech}(\mathbf{x}_t^* \mathbf{x}_t^{*\prime}) = \mathbf{D}_m^+(\Sigma_{\mathbf{x}}^{-1/2} \otimes \Sigma_{\mathbf{x}}^{-1/2})(\mathbf{x}_t \otimes \mathbf{x}_t)$. Since $\mathbf{x}_t \otimes \mathbf{x}_t = \mathrm{vec}(\mathbf{x}_t \mathbf{x}_t') = \mathbf{D}_m \mathrm{vech}(\mathbf{x}_t \mathbf{x}_t') = \mathbf{D}_m \mathbf{y}_t$, it follows that $\mathbf{y}_t^* = \mathbf{R}\mathbf{y}_t$, where $\mathbf{R} = \mathbf{D}_m^+(\Sigma_{\mathbf{x}}^{-1/2} \otimes \Sigma_{\mathbf{x}}^{-1/2})\mathbf{D}_m$. Then we have $\Sigma_{\mathbf{y}^*} = \mathbf{R}\Sigma_{\mathbf{y}}\mathbf{R}'$, where $\Sigma_{\mathbf{y}}$ is given by Theorem 3.

It is immediate to see that this concept extends the corresponding kurtosis measure known for univariate time series (Hamilton, 1994, p. 746). In fact, if $x = (x_t)$ is univariate $(m = 1)$, then $k(x) = E(x_t^{*4}) = E(\sigma_{\mathbf{x}}^{-2} x_t^4) = E(x_t^4)/[\mathrm{var}(x_t)]^2$. Furthermore, the above definition relates with the multivariate kurtosis measure proposed by Mardia (1970). In the example of Section 3, both measures will be reported and compared. Given a multivariate process $\mathbf{x} = (\mathbf{x}_t)$, Mardia's kurtosis of $\mathbf{x}$ is defined to be $\beta(\mathbf{x}) = E\{[(\mathbf{x}_t - \boldsymbol{\mu}_{\mathbf{x}})'\Sigma_{\mathbf{x}}^{-1}(\mathbf{x}_t - \boldsymbol{\mu}_{\mathbf{x}})]^2\} \in \mathbb{R}$, where $\boldsymbol{\mu}_{\mathbf{x}}$ and $\Sigma_{\mathbf{x}}$ are the mean and the covariance matrix of $\mathbf{x}$, respectively. One disadvantage of Mardia's definition depends on the fact that not all order mixed moments are taken in account. So our different approach constitutes a merit of the new definition with respect to that of Mardia.

Lemma. *The above defined kurtosis measure for multivariate MS GARCH models is invariant under nonsingular time-independent linear transformation of the random vector* $\mathbf{x}_t$.

For a random sample of size T drawn from $\mathbf{x}$, the measure of the sample kurtosis corresponding to $\mathbf{k}(\mathbf{x})$ is given by

$$\widehat{\mathbf{k}}_T(\mathbf{x}) = \mathbf{S}_{\mathbf{x}}^{-1/2}\left(T^{-1}\sum_{t=1}^{T}\mathbf{y}_t^* \otimes \mathbf{y}_t^*\right),$$

where $\mathbf{S}_{\mathbf{x}}$ is the sample covariance matrix.

Theorem 6. *Under Assumptions 1–4 and 6, the* K^2*-dimensional sample estimator* $\widehat{\mathbf{k}}_T(\mathbf{x})$ *is consistent, that is, it converges to* $\mathbf{k}(\mathbf{x})$ *with probability 1 as T goes to infinity.*

In the univariate case, Bauwens, Preminger, and Rombouts (2010, theorem 2.1) provide sufficient conditions for the geometric ergodicity of the extended process $Z_t = (x_t \quad h_{t+1} \quad s_{t+1})'$. The geometric ergodicity ensures not only that a unique stationary probability measure for the process exists, but also that the chain, irrespective of its initialization, converges to it at a geometric rate with respect to the total variation norm. Markov chains with this property satisfy the central limit theorem for any given starting value given the existence of suitable moments (Jones, 2004, theorem 5). In this case, the above-defined sample kurtosis is also asymptotically normally distributed. An open problem is the proof of the geometric ergodicity in the multivariate MS case.

In empirical applications, it might be useful to compare the above-defined measure of multivariate kurtosis with that of a centered normal distribution in order to investigate whether an MS GARCH process has close features to normality or not. Recall that the fourth moment of an m-dimensional normal distributed process with covariance matrix Ω is given by (see the Supplementary Material, Section 5)

$$(\mathbf{K}_{mm} \otimes \mathbf{I}_{m^2} + \mathbf{I}_{m^4})\mathrm{vec}(\Omega \otimes \Omega) + \mathrm{vec}(\Omega) \otimes \mathrm{vec}(\Omega). \tag{6}$$

To obtain the multivariate kurtosis measure, as defined above, we must premultiply the last $m^4 \times 1$ vector by the $K^2 \times m^4$ matrix $\mathbf{D}_m^+ \otimes \mathbf{D}_m^+$.

2.3 Impulse-Response Functions for the Volatility

Following Hafner (2003, p. 34), an impulse-response function for the volatility is given by the difference of a "shock scenario" and a "baseline scenario"

$$\mathbf{v}_\tau(\boldsymbol{\eta}_t) = E\big(\mathbf{h}_{t+\tau}|\boldsymbol{\eta}_t, \mathcal{F}_{t-1}\big) - E(\mathbf{h}_{t+\tau}|\mathcal{F}_{t-1})$$

for $\tau \geq 0$. Let $\mathcal{V}_\tau$ be the $(Mn) \times 1$ matrix whose i-th block is given by

$$\pi_i\big[E(\mathbf{z}_{t+\tau}|\boldsymbol{\eta}_t, \mathcal{F}_{t-1}, s_{t+\tau} = i) - E(\mathbf{z}_{t+\tau}|\mathcal{F}_{t-1}, s_{t+\tau} = i)\big]$$

for $i = 1, \ldots, M$ and $\tau \geq 0$. We set $\mathcal{V} = \mathcal{V}_0$. The following matrix relations are proved in the Supplementary Material, Sections 6 and 7:

$$\mathcal{V} = \left[\mathbf{I}_M \otimes \left(\mathbf{g}\mathbf{D}_m^+\right)\right]\mathcal{H}\mathbf{u}_t, \tag{7}$$

$$\mathcal{V}_\tau = \mathbf{P}(\Phi)\mathcal{V}_{\tau-1} = [\mathbf{P}(\Phi)]^\tau \mathcal{V}, \tag{8}$$

where $\mathcal{H}$ is the $(Mm^2) \times m^2$ matrix whose i-th block is given by

$$\pi_i E\left(\mathbf{H}_t^{1/2} \otimes \mathbf{H}_t^{1/2} | \mathcal{F}_{t-1}, s_t = i\right) \in \mathbb{R}^{m^2 \times m^2}$$

for $i = 1, \ldots, M$. Then we have:

Theorem 7. *Under Assumptions 1–5, the impulse-response functions for the volatility of the multivariate* MS GARCH(p, q) *model in* Equations (1) *and* (2) *are given by the following matrix expression in closed form:*

$$\mathbf{v}_\tau(\boldsymbol{\eta}_t) = (\mathbf{e}' \otimes \mathbf{f}')[\mathbf{P}(\Phi)]^\tau \left[\mathbf{I}_M \otimes \left(\mathbf{g}\mathbf{D}_m^+\right)\right]\mathcal{H}\mathbf{u}_t.$$

As a consequence, the long-run effect is given by

$$\sum_{\tau=0}^{\infty} \mathbf{v}_\tau(\boldsymbol{\eta}_t) = (\mathbf{e}' \otimes \mathbf{f}')[\mathbf{I}_{Mn} - \mathbf{P}(\Phi)]^{-1} \left[\mathbf{I}_M \otimes \left(\mathbf{g}\mathbf{D}_m^+\right)\right]\mathcal{H}\mathbf{u}_t.$$

For a multivariate state-invariant GARCH model (case $M = 1$), the formula in Theorem 7 becomes

$$\mathbf{v}_\tau(\boldsymbol{\eta}_t) = \phi_\tau \mathbf{L}_m \left(\mathbf{H}_t^{1/2} \otimes \mathbf{H}_t^{1/2}\right)\mathbf{D}_m \mathrm{vech}\left(\boldsymbol{\eta}_t\boldsymbol{\eta}_t' - \mathbf{I}_{m^2}\right), \tag{9}$$

which is formula (24) in Hafner (2003), where $\phi_\tau = \mathbf{f}'\Phi^\tau\mathbf{g} \in \mathbb{R}^{K \times K}$, and Φ is the $n \times n$ matrix obtained by setting $s_t = 1$ in the definition of Φ_t. In this case, we have $E(\mathbf{H}_t^{1/2} \otimes \mathbf{H}_t^{1/2} | \mathcal{F}_{t-1}, s_t = 1) = E(\mathbf{H}_t^{1/2} \otimes \mathbf{H}_t^{1/2} | \mathcal{F}_{t-1}) = \mathbf{H}_t^{1/2} \otimes \mathbf{H}_t^{1/2}$. This also gives the VMA(∞) representation $\mathbf{y}_t = \sigma + \sum_{i=0}^{\infty} \phi_i \boldsymbol{\varepsilon}_{t-i}$, where $\sigma = (\mathbf{I}_K - \sum_{i=1}^{r} \mathbf{a}_i)^{-1}\mathbf{c}$ and $\phi_i = \mathbf{f}'\Phi^i\mathbf{g}$ and $\mathbf{a}_i$ denotes $\mathbf{a}_i(s_t)$ for the unique regime $s_t = 1$. This reproduces formulas 5–7 in Hafner (2003) in the single regime case.

3 An Empirical Example

In this section, we consider an MS(2) GARCH(1, 1) model from Lee and Yoder (2007, p. 1258) in order to illustrate our theoretical results. These authors use bivariate two-state MS GARCH models to estimate the hedge ratios through the second moments of futures and spot prices in the corn and nickel commodity markets. The parameter estimates of their MS BEKK-GARCH model (as named after Baba, Engle, Kraft and Kroner specification) for corn futures contracts (traded on the Chicago Board of Trade; sample period from January 2, 1991 to December 31, 2003) are presented in Table 2 of Lee and Yoder (2007). This model can be nested into model (1) and (2) as the estimated conditional mean is close to zero. So we derive the following MS(M) multivariate GARCH(p, q) model such that $M =$

$m = 2$ and $p = q = 1$ (hence, $K = 3$ and $n = 6$), that is, a two-state bivariate GARCH(1, 1) model:

$$\mathbf{x}_t = \mathbf{H}_t^{1/2}\boldsymbol{\eta}_t \qquad \boldsymbol{\eta}_t \sim \mathrm{NID}(\boldsymbol{0}, \mathbf{I}_2)$$
$$\mathbf{h}_t = \mathbf{c}(s_t) + \mathbf{A}(s_t)\mathbf{y}_{t-1} + \mathbf{B}(s_t)\mathbf{h}_{t-1},$$

where $\mathbf{x}_t = (r_{c,t} \quad r_{f,t})' \in \mathbb{R}^2$, with $r_{c,t}$ and $r_{f,t}$ returns on the spot and futures at time t, respectively, $\mathbf{y}_t = \mathrm{vech}(\mathbf{x}_t\mathbf{x}_t') \in \mathbb{R}^3$ and $\mathbf{h}_t = \mathrm{vech}(\mathbf{H}_t) \in \mathbb{R}^3$,

$$\mathbf{c}(s_t) = (3.4501 - 1.7193s_t \qquad 5.6368 - 2.8120s_t \qquad 5.8298 - 2.9078s_t)',$$

$$\mathbf{A}(s_t) = \begin{pmatrix} 0.2523s_t - 0.0846 & -0.5176s_t + 0.6362 & 0.2952s_t - 0.2113 \\ 0.2804s_t - 0.1618 & -0.2075s_t + 0.1961 & 0.2161s_t - 0.2242 \\ 0.2952s_t - 0.2113 & -0.1999s_t + 0.1918 & 0.1133s_t - 0.1125 \end{pmatrix},$$

$$\mathbf{B}(s_t) = \begin{pmatrix} -0.1435s_t + 0.5511 & 0.6451s_t - 1.1356 & -0.5714s_t + 1.1617 \\ -0.4768s_t + 0.6197 & 0.9397s_t - 1.2139 & -0.1263s_t + 0.4563 \\ 0.0381s_t + 0.012 & -0.3438s_t + 0.2476 & 1.0100s_t - 0.8255 \end{pmatrix},$$

and $s_t \in \{1, 2\}$. The transition probability matrix is

$$\mathbf{P} = \begin{pmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{pmatrix},$$

where $p_{11} = 0.6743$ and $p_{22} = 0.5349$. The unconditional probabilities are $\pi_1 = 0.5881$ and $\pi_2 = 0.4119$. Then the process $\mathbf{x} = (\mathbf{x}_t)$ is obtained from the above model by using estimated parameters. The basic assumption that the eigenvalues of the 12×12 matrix $\mathbf{P}(\Phi)$ are less than 1 in modulus is fulfilled because its spectral radius is 0.8414. Then the second moments of $\mathbf{x}$ are finite. Applying the formula in Theorem 1 gives

$$\boldsymbol{\sigma}_{\mathbf{x}} = (6.8757 \qquad 5.1479 \qquad 5.8749)' \in \mathbb{R}^3,$$

hence the unconditional covariance matrix is

$$\Sigma_{\mathbf{x}} = \mathrm{var}(\mathbf{x}_t) = \begin{pmatrix} 6.8757 & 5.1479 \\ 5.1479 & 5.8749 \end{pmatrix}.$$

The eigenvalues of $\Sigma_{\mathbf{x}}$ are 1.2032 and 11.5475. Assumption 5 is satisfied as the spectral radius of the 72×72 matrix $\mathbf{P}(\Phi \otimes \Phi)$ is 0.8264, and the rank of the 72×18 matrix $\mathbf{Q}$ is 18. Applying the formula in Theorem 3, the unconditional fourth moments of $\mathbf{x}$ are given by $\mathrm{vech}(\Sigma_{\mathbf{y}})$, where

$$\Sigma_{\mathbf{y}} = \begin{pmatrix} 52.6713 & 46.3358 & 25.9558 \\ 46.3358 & 45.0175 & 21.6136 \\ 25.9558 & 21.6136 & 16.2993 \end{pmatrix}.$$

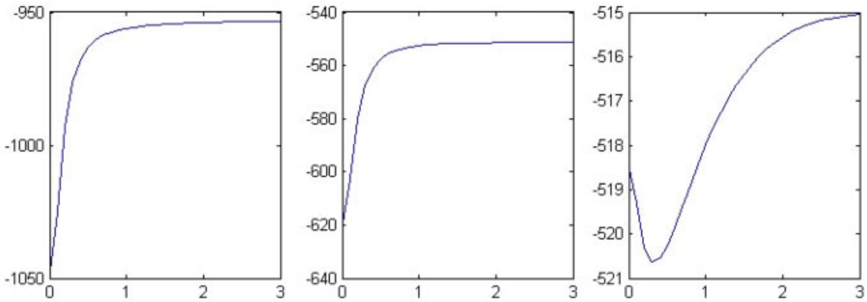

Figure 1 Spectral density functions of the squares $r_{c,t}^2$ (left panel), $r_{f,t}^2$ (middle panel), and cross product $r_{c,t}r_{f,t}$ (right panel) obtained from the bivariate MS(2) GARCH(1, 1) model in Lee and Yoder (2007) and described in Section 3.

The eigenvalues of $\Sigma_{\mathbf{y}}$ are 1.3534, 4.8533, and 107.7814. The kurtosis of $\mathbf{x}$ is the 9×1 vector $\mathbf{k}(\mathbf{x}) = \text{vec}\Sigma_{\mathbf{y}^*}$, where

$$\Sigma_{\mathbf{y}^*} = \begin{pmatrix} 5.6369 & 16.5031 & 17.3930 \\ 16.5031 & 17.4190 & 18.3599 \\ 17.3930 & 18.3599 & 19.3531 \end{pmatrix}.$$

Computing the value of Mardia's kurtosis for this exercise gives 5.2856, which is sufficiently close to the first element of the kurtosis vector $\mathbf{k}(\mathbf{x})$ in our definition. This further confims that not all order mixed moments are taken in account.

Using the estimated parameters, we are able to reproduce the empirical kurtosis of the single returns of corn reported in Table 1 in Lee and Yoder (2007), which is 14.3795. Moreover, by Theorem 5, we derive the spectral density matrix of $\mathbf{y}$, and plot the spectral density functions of the squares $r_{c,t}^2$, $r_{f,t}^2$, and cross product $r_{c,t}r_{f,t}$ in Figure 1. The curves of the squares increase quickly to high values at high frequency, clearly indicating short-term movements in returns. However, it is worth to notice that the cross product exhibits less short-term interaction in favour of more important medium and low frequencies, indicating delayed effects. Finally, we complete the analysis with impulse-response functions for the volatility using the expression in Theorem 7. As mentioned before, the impulse-response function for the volatility is given by the difference of a "shock scenario" and a "baseline scenario." In Figure 2, we report impulse-responses for the volatility $\mathbf{h}$, which correspond (from left to right) to spot returns, mixed term, and future returns. The impulse-responses show an increase (with different magnitude in the three cases) over the first 30 days and a decrease thereafter.

Further applications and illustrative examples with univariate MS GARCH models from Bauwens, Preminger, and Rombouts (2010, pp. 229 and 235) and Zhang, Li, and Yuen (2006, p. 592) are provided in the Supplementary Material, Section 9.

4 Conclusion

We provide matrix expressions in closed form for the second and fourth unconditional moments of multivariate MS GARCH processes in their general vector specification. Unlike

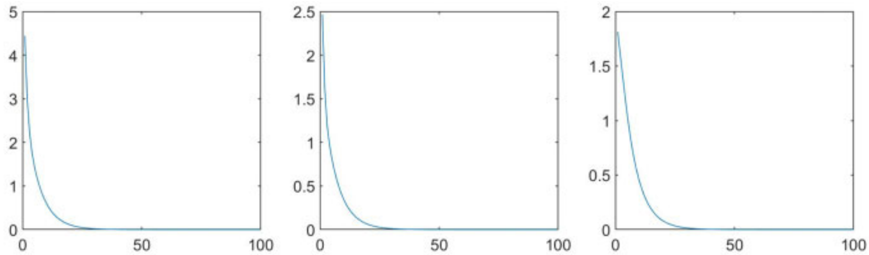

Figure 2 Impulse-response functions for the volatility obtained from the bivariate MS(2) GARCH(1, 1) model in Lee and Yoder (2007) and described in Section 3. They correspond (from left to right) to spot returns, mixed term, and future returns.

second moments, fourth moments crucially depend on the innovation distribution, which we assume to be spherical distributions in the sense of Hafner (2003). We propose some potential applications of the obtained results, such as the derivation of the spectral density matrix of the squares and cross products, a new definition of the multivariate kurtosis measure to recognize heavy-tailed features in real data, and a matrix expression in closed form of the impulse-response function for the volatility. These theoretical results are illustrated with an empirical example in the article, while other applications and examples are given in the Supplementary Material. Further analyses are possible, such as, for example, the issue of the temporal aggregation of multivariate MS GARCH models, the analysis of the multivariate kurtosis as a function of the sampling frequency, and the estimation of the asymmetric volatility under skewed distributions.

Acknowledgments

This work is financially supported by FAR (2017) research grant of the University of Modena and Reggio Emilia, Italy. We would like to thank the Editor in Chief of the journal, Professor Allan Timmermann, the Associate Editor, and the two anonymous referees for their constructive comments and very useful suggestions and remarks that were most valuable for improvement of the final version of the paper.

Supplementary Data

Supplementary data are available at *Journal of Financial Econometrics* online.

Appendix

Proof of Theorem 1. Let $\mathbf{U}_t$ be the (Mn)-dimensional vector, whose i-th block is the n-dimensional vector $\pi_i E(\mathbf{z}_t|s_t = i)$ for $i = 1, \ldots, M$, that is,

$$\mathbf{U}_t = \left(\pi_1 E\left(\mathbf{z}_t'|s_t = 1\right) \cdots \pi_M E\left(\mathbf{z}_t'|s_t = M\right)\right)' \in \mathbb{R}^{Mn}.$$

Starting from Equation (4), for $i = 1, \ldots, M$, we have

$$\begin{aligned}\pi_i E(\mathbf{z}_t|s_t = i) &= \pi_i E\big(\Phi(i)\mathbf{z}_{t-1} + \mathbf{c}(i) + \Sigma_t|s_t = i\big)\\ &= \pi_i \Phi(i) E(\mathbf{z}_{t-1}|s_t = i) + \pi_i \mathbf{c}(i) + \pi_i E(\Sigma_t|s_t = i)\\ &= \pi_i \Phi(i) E(\mathbf{z}_{t-1}|s_t = i) + \pi_i \mathbf{c}(i)\\ &= \sum_{j=1}^{M} p_{ji} \Phi(i) \pi_j E(\mathbf{z}_{t-1}|s_{t-1} = j) + \pi_i \mathbf{c}(i)\end{aligned}$$

as $E(\Sigma_t|s_t = i) = \mathbf{g}E\big[E(\boldsymbol{\sigma}_t|\mathcal{F}_{t-1}, s_t = i)E(\mathbf{u}_t)\big] = \mathbf{0}$. For $i = 1, \ldots, M$, let $\mathbf{U}_{it}$ and $\mathbf{c}_i$ be the i-th block of $\mathbf{U}_t$ and $\mathbf{c}$, respectively. For $i, j = 1, \ldots, M$, let $\mathbf{P}(\Phi)_{ij}$ be the (i, j) block of $\mathbf{P}(\Phi)$. Thus, $\mathbf{U}_{it}$ and $\mathbf{c}_i$ are $n \times 1$, and $\mathbf{P}(\Phi)_{ij}$ is $n \times n$. Using these notations, the last formula becomes

$$\mathbf{U}_{it} = \sum_{j=1}^{M} \mathbf{P}(\Phi)_{ij} \mathbf{U}_{jt} + \mathbf{c}_i$$

for $i = 1, \ldots, M$. Then the matrix relation

$$\mathbf{U}_t = \mathbf{P}(\Phi)\mathbf{U}_t + \mathbf{c}$$

holds. If $\rho(\mathbf{P}(\Phi)) < 1$, where $\rho(\cdot)$ denotes the spectral radius, the matrix $\mathbf{I}_{Mn} - \mathbf{P}(\Phi)$ is invertible. Then $\mathbf{U}_t$ can be expressed in closed form as

$$\mathbf{U}_t = (\mathbf{I}_{Mn} - \mathbf{P}(\Phi))^{-1}\mathbf{c}.$$

So $\mathbf{U}_t$ is time-invariant, and we set $\mathbf{U} = \mathbf{U}_t$ for every $t \in \mathbb{Z}$. The unconditional mean of $\mathbf{z}_t$ is given by $E(\mathbf{z}_t) = (\mathbf{e}' \otimes \mathbf{I}_n)\mathbf{U}$. By construction, it follows that $E(\mathbf{y}_t) = (\mathbf{e}' \otimes \mathbf{f}')\mathbf{U}$. So the process $(\mathbf{z}_t)$ [and hence $(\mathbf{y}_t)$] is first-order stationary. Finally, we have

$$\boldsymbol{\sigma}_{\mathbf{x}} = \mathrm{vech}(\Sigma_{\mathbf{x}}) = E\big(\mathrm{vech}\big(\mathbf{x}_t\mathbf{x}_t'\big)\big) = E(\mathbf{y}_t) = (\mathbf{e}' \otimes \mathbf{f}')\mathbf{U},$$

where $\mathbf{U} = (\mathbf{I}_{Mn} - \mathbf{P}(\Phi))^{-1}\mathbf{c}$. This completes the proof.

Proof of Theorem 2. We have

$$\begin{aligned}\mathrm{var}\big[\mathrm{vec}\big(\mathbf{x}_t\mathbf{x}_t'\big)|\mathcal{F}_{t-1}, \Delta_t\big] &= \mathrm{var}(\mathbf{x}_t \otimes \mathbf{x}_t|\mathcal{F}_{t-1}, \Delta_t)\\ &= \mathrm{var}\Big[\Big(\mathbf{H}_t^{1/2} \otimes \mathbf{H}_t^{1/2}\Big)(\boldsymbol{\eta}_t \otimes \boldsymbol{\eta}_t)|\mathcal{F}_{t-1}, \Delta_t\Big]\\ &= \Big(\mathbf{H}_t^{1/2} \otimes \mathbf{H}_t^{1/2}\Big)\mathrm{var}(\boldsymbol{\eta}_t \otimes \boldsymbol{\eta}_t)\Big(\mathbf{H}_t^{1/2} \otimes \mathbf{H}_t^{1/2}\Big)\\ &= 2c\Big(\mathbf{H}_t^{1/2} \otimes \mathbf{H}_t^{1/2}\Big)\mathbf{D}_m\mathbf{D}_m^+\Big(\mathbf{H}_t^{1/2} \otimes \mathbf{H}_t^{1/2}\Big)\\ &\quad + (c-1)\Big(\mathbf{H}_t^{1/2} \otimes \mathbf{H}_t^{1/2}\Big)\mathrm{vec}(\mathbf{I}_m)\mathrm{vec}(\mathbf{I}_m)'\Big(\mathbf{H}_t^{1/2} \otimes \mathbf{H}_t^{1/2}\Big)\\ &= 2c\Big(\mathbf{H}_t^{1/2} \otimes \mathbf{H}_t^{1/2}\Big)\mathbf{D}_m\mathbf{D}_m^+\Big(\mathbf{H}_t^{1/2} \otimes \mathbf{H}_t^{1/2}\Big) + (c-1)\mathrm{vec}(\mathbf{H}_t)\mathrm{vec}(\mathbf{H}_t)'\end{aligned}$$

as $\mathrm{var}(\boldsymbol{\eta}_t \otimes \boldsymbol{\eta}_t) = 2c\mathbf{D}_m\mathbf{D}_m^+ + (c-1)\mathrm{vec}(\mathbf{I}_m)\mathrm{vec}(\mathbf{I}_m)'$ by Lemma 3 of Hafner (2003). Theorem 3.11 of Magnus (1988) implies

$$\Big(\mathbf{H}_t^{1/2} \otimes \mathbf{H}_t^{1/2}\Big)\mathbf{D}_m\mathbf{D}_m^+ = \mathbf{D}_m\mathbf{D}_m^+\Big(\mathbf{H}_t^{1/2} \otimes \mathbf{H}_t^{1/2}\Big),$$

hence

$$\mathrm{var}(\mathbf{x}_t \otimes \mathbf{x}_t|\mathcal{F}_{t-1}, \Delta_t) = 2c\mathbf{D}_m\mathbf{D}_m^+(\mathbf{H}_t \otimes \mathbf{H}_t) + (c-1)\mathrm{vec}(\mathbf{H}_t)\,\mathrm{vec}(\mathbf{H}_t)'.$$

By definition, $\mathbf{y}_t = \mathrm{vech}(\mathbf{x}_t\mathbf{x}_t') = \mathbf{D}_m^+\,\mathrm{vec}(\mathbf{x}_t\mathbf{x}_t') = \mathbf{L}_m\,\mathrm{vec}(\mathbf{x}_t\mathbf{x}_t')$. In fact, one could replace $\mathbf{D}_m^+$ by $\mathbf{L}_m$ by the symmetry of $\mathbf{x}_t\mathbf{x}_t'$. This gives

$$\begin{aligned}\mathrm{var}(\mathbf{y}_t|\mathcal{F}_{t-1}, \Delta_t) &= \mathrm{var}(\mathbf{L}_m\mathbf{x}_t \otimes \mathbf{x}_t|\mathcal{F}_{t-1}, \Delta_t) = \mathbf{L}_m\mathrm{var}(\mathbf{x}_t \otimes \mathbf{x}_t|\mathcal{F}_{t-1}, \Delta_t)\mathbf{L}_m' \\ &= 2c\mathbf{L}_m\mathbf{D}_m\mathbf{D}_m^+(\mathbf{H}_t \otimes \mathbf{H}_t)\mathbf{L}_m' + (c-1)\mathbf{L}_m\,\mathrm{vec}(\mathbf{H}_t)\,\mathrm{vec}(\mathbf{H}_t)'\mathbf{L}_m' \\ &= 2c\mathbf{D}_m^+(\mathbf{H}_t \otimes \mathbf{H}_t)\mathbf{L}_m' + (c-1)\mathbf{L}_m\,\mathrm{vec}(\mathbf{H}_t)[\mathbf{L}_m\,\mathrm{vec}(\mathbf{H}_t)]'\end{aligned}$$

as $\mathbf{L}_m\mathbf{D}_m = \mathbf{I}_K$ by Lütkepohl (2007, p. 664). Applying the vec operator yields

$$\begin{aligned}\mathrm{vec}[\mathrm{var}(\mathbf{y}_t|\mathcal{F}_{t-1}, \Delta_t)] &= 2c(\mathbf{L}_m \otimes \mathbf{D}_m^+)\,\mathrm{vec}(\mathbf{H}_t \otimes \mathbf{H}_t) \\ &\quad + (c-1)[\mathbf{L}_m\,\mathrm{vec}(\mathbf{H}_t)] \otimes [\mathbf{L}_m\,\mathrm{vec}(\mathbf{H}_t)] \\ &= 2c(\mathbf{L}_m \otimes \mathbf{D}_m^+)(\mathbf{I}_m \otimes \mathbf{K}_{mm} \otimes \mathbf{I}_m)[\mathrm{vec}(\mathbf{H}_t) \otimes \mathrm{vec}(\mathbf{H}_t)] \\ &\quad + (c-1)(\mathbf{L}_m \otimes \mathbf{L}_m)[\mathrm{vec}(\mathbf{H}_t) \otimes \mathrm{vec}(\mathbf{H}_t)] \\ &= [2c(\mathbf{L}_m \otimes \mathbf{D}_m^+)(\mathbf{I}_m \otimes \mathbf{K}_{mm} \otimes \mathbf{I}_m) + (c-1)(\mathbf{L}_m \otimes \mathbf{L}_m)][\mathrm{vec}(\mathbf{H}_t) \otimes \mathrm{vec}(\mathbf{H}_t)].\end{aligned}$$

By definition, $\mathbf{D}_m\mathbf{h}_t = \mathbf{D}_m\,\mathrm{vech}(\mathbf{H}_t) = \mathrm{vec}(\mathbf{H}_t)$. Using this relation, we get

$$\begin{aligned}\mathrm{vec}[\mathrm{var}(\mathbf{y}_t|\mathcal{F}_{t-1}, \Delta_t)] = {} & [2c(\mathbf{L}_m \otimes \mathbf{D}_m^+)(\mathbf{I}_m \otimes \mathbf{K}_{mm} \otimes \mathbf{I}_m) + (c-1)(\mathbf{L}_m \otimes \mathbf{L}_m)] \\ & \times (\mathbf{D}_m \otimes \mathbf{D}_m)(\mathbf{h}_t \otimes \mathbf{h}_t).\end{aligned}$$

But

$$\begin{aligned}\mathrm{var}(\mathbf{y}_t|\mathcal{F}_{t-1}, \Delta_t) &= E(\mathbf{y}_t\mathbf{y}_t'|\mathcal{F}_{t-1}, \Delta_t) - E(\mathbf{y}_t|\mathcal{F}_{t-1}, \Delta_t)E(\mathbf{y}_t'|\mathcal{F}_{t-1}, \Delta_t) \\ &= E(\mathbf{y}_t\mathbf{y}_t'|\mathcal{F}_{t-1}, \Delta_t) - E(\mathbf{h}_t|\mathcal{F}_{t-1}, \Delta_t)E(\mathbf{h}_t'|\mathcal{F}_{t-1}, \Delta_t) = E(\mathbf{y}_t\mathbf{y}_t'|\mathcal{F}_{t-1}, \Delta_t) - \mathbf{h}_t\mathbf{h}_t'.\end{aligned}$$

Hence,

$$\begin{aligned}\boldsymbol{\Sigma}_\epsilon &= E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = E[E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'|\mathcal{F}_{t-1}, \Delta_t)] = E[\mathrm{var}(\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1}, \Delta_t)] \\ &= E[\mathrm{var}(\mathbf{y}_t|\mathcal{F}_{t-1}, \Delta_t)] = E[E(\mathbf{y}_t\mathbf{y}_t'|\mathcal{F}_{t-1}, \Delta_t)] - E(\mathbf{h}_t\mathbf{h}_t') \\ &= E(\mathbf{y}_t\mathbf{y}_t') - E(\mathbf{h}_t\mathbf{h}_t') = \boldsymbol{\Sigma}_\mathbf{y} - \boldsymbol{\Sigma}_\mathbf{h}\end{aligned}$$

by using Equation (3). This proves the first relation in the statement of Theorem 2. From above, we can also write

$$\begin{aligned}\mathrm{vec}\boldsymbol{\Sigma}_\mathbf{y} &= E[\mathrm{vec}\,\mathrm{var}(\mathbf{y}_t|\mathcal{F}_{t-1}, \Delta_t)] + \mathrm{vec}\boldsymbol{\Sigma}_\mathbf{h} \\ &= [2c(\mathbf{L}_m \otimes \mathbf{D}_m^+)(\mathbf{I}_m \otimes \mathbf{K}_{mm} \otimes \mathbf{I}_m) + (c-1)(\mathbf{L}_m \otimes \mathbf{L}_m)] \\ &\quad \times (\mathbf{D}_m \otimes \mathbf{D}_m)\,\mathrm{vec}\boldsymbol{\Sigma}_\mathbf{h} + \mathrm{vec}\boldsymbol{\Sigma}_\mathbf{h} \\ &= [2c(\mathbf{L}_m \otimes \mathbf{D}_m^+)(\mathbf{I}_m \otimes \mathbf{K}_{mm} \otimes \mathbf{I}_m)(\mathbf{D}_m \otimes \mathbf{D}_m) + (c-1)(\mathbf{L}_m \otimes \mathbf{L}_m)(\mathbf{D}_m \otimes \mathbf{D}_m) + \mathbf{I}_{K^2}]\,\mathrm{vec}\boldsymbol{\Sigma}_\mathbf{h} \\ &= [2c(\mathbf{L}_m \otimes \mathbf{D}_m^+)(\mathbf{I}_m \otimes \mathbf{K}_{mm} \otimes \mathbf{I}_m)(\mathbf{D}_m \otimes \mathbf{D}_m) + (c-1)(\mathbf{L}_m\mathbf{D}_m) \otimes (\mathbf{L}_m\mathbf{D}_m) + \mathbf{I}_{K^2}]\,\mathrm{vec}\boldsymbol{\Sigma}_\mathbf{h} \\ &= [2c(\mathbf{L}_m \otimes \mathbf{D}_m^+)(\mathbf{I}_m \otimes \mathbf{K}_{mm} \otimes \mathbf{I}_m)(\mathbf{D}_m \otimes \mathbf{D}_m) + (c-1)\mathbf{I}_K \otimes \mathbf{I}_K + \mathbf{I}_{K^2}]\,\mathrm{vec}\boldsymbol{\Sigma}_\mathbf{h} \\ &= \mathbf{G}_K\,\mathrm{vec}\boldsymbol{\Sigma}_\mathbf{h}\end{aligned}$$

as $\mathbf{L}_m\mathbf{D}_m = \mathbf{I}_K$ and $\mathbf{I}_K \otimes \mathbf{I}_K = \mathbf{I}_{K^2}$. Here, $\mathbf{G}_K$ is the matrix defined in the statement of Theorem 2. Finally, we have

$$\mathrm{vec}\Sigma_{\varepsilon} = \mathrm{vec}\Sigma_{\mathbf{y}} - \mathrm{vec}\Sigma_{\mathbf{h}} = (\mathbf{G}_K - \mathbf{I}_{K^2})\,\mathrm{vec}\Sigma_{\mathbf{h}}$$

as required. An alternative proof of the relation

$$\mathrm{vec}\Sigma_{\varepsilon} = \mathrm{vec}\Sigma_{\mathbf{y}} - \mathrm{vec}\Sigma_{\mathbf{h}}$$

can also be obtained from equation $\boldsymbol{\varepsilon}_t = \mathbf{y}_t - \mathbf{h}_t$. In fact, we have

$$\begin{aligned}\mathrm{vec}\Sigma_{\varepsilon} &= E\big[(\mathbf{y}_t - \mathbf{h}_t) \otimes (\mathbf{y}_t - \mathbf{h}_t)\big]\\ &= \mathrm{vec}\Sigma_{\mathbf{y}} - E(\mathbf{y}_t \otimes \mathbf{h}_t) - E(\mathbf{h}_t \otimes \mathbf{y}_t) + \mathrm{vec}\Sigma_{\mathbf{h}}.\end{aligned}$$

But one can directly prove that

$$\begin{aligned}\mathrm{vec}\Sigma_{\mathbf{h}} &= E(\mathbf{y}_t \otimes \mathbf{h}_t) = E(\mathbf{h}_t \otimes \mathbf{y}_t)\\ &= (\mathbf{D}_m^+ \otimes \mathbf{D}_m^+)E\left[\left(\mathbf{H}_t^{1/2}\right)^{\otimes 4}\right][\mathrm{vec}(\mathbf{I}_m) \otimes \mathrm{vec}(\mathbf{I}_m)],\end{aligned}$$

where $(\mathbf{H}_t^{1/2})^{\otimes 4}$ denotes the Kronecker product of four copies of $\mathbf{H}_t^{1/2}$.

Proof of Theorem 3. Let $\mathbf{W}_t$ and $\mathbf{V}_t$ be defined as $\mathbf{U}$ by replacing $\mathbf{z}_t$ by $\mathbf{z}_t \otimes \mathbf{z}_t$ and $\mathbf{h}_t \otimes \mathbf{h}_t$, respectively. Let us consider the $(Mn^2) \times 1$ matrix defined by

$$\mathbf{S}_t = \begin{pmatrix} \pi_1 E(\Sigma_t \otimes \Sigma_t | s_t = 1) \\ \pi_2 E(\Sigma_t \otimes \Sigma_t | s_t = 2) \\ \vdots \\ \pi_M E(\Sigma_t \otimes \Sigma_t | s_t = M) \end{pmatrix}.$$

Since $\Sigma_t = \mathbf{g}\boldsymbol{\varepsilon}_t$ and $\mathbf{z}_t = \mathbf{f}\mathbf{y}_t$, Theorem 2 implies

$$\begin{aligned}E(\Sigma_t \otimes \Sigma_t) &= (\mathbf{g} \otimes \mathbf{g})\,\mathrm{vec}\Sigma_{\varepsilon} = (\mathbf{g} \otimes \mathbf{g})(\mathbf{G}_K - \mathbf{I}_{K^2})\,\mathrm{vec}\Sigma_{\mathbf{h}} = \tilde{\mathbf{g}}\,\mathrm{vec}\Sigma_{\mathbf{h}}\\ E(\mathbf{z}_t \otimes \mathbf{z}_t) &= (\mathbf{f} \otimes \mathbf{f})\,\mathrm{vec}\Sigma_{\mathbf{y}} = (\mathbf{f} \otimes \mathbf{f})\mathbf{G}_K\,\mathrm{vec}\Sigma_{\mathbf{h}} = \tilde{\mathbf{f}}\,\mathrm{vec}\Sigma_{\mathbf{h}}.\end{aligned}$$

Then we have $\mathbf{S}_t = (\mathbf{I}_M \otimes \tilde{\mathbf{g}})\mathbf{V}_t$ and $\mathbf{W}_t = (\mathbf{I}_M \otimes \tilde{\mathbf{f}})\mathbf{V}_t$. Now we are going to prove the following matrix relation

$$\mathbf{W}_t = \mathbf{P}(\Phi \otimes \Phi)\mathbf{W}_t + \mathbf{D}\mathbf{U} + \mathbf{C} + \mathbf{S}_t.$$

Starting from Equation (4), for $i = 1, \ldots, M$, we have

$$\begin{aligned}\pi_i E(\mathbf{z}_t \otimes \mathbf{z}_t | s_t = i) &= \pi_i E\big[(\Phi(i)\mathbf{z}_{t-1} + \mathbf{c}(i) + \Sigma_t) \otimes (\Phi(i)\mathbf{z}_{t-1} + \mathbf{c}(i) + \Sigma_t) | s_t = i\big]\\ &= \pi_i[\Phi(i) \otimes \Phi(i)]E(\mathbf{z}_{t-1} \otimes \mathbf{z}_{t-1} | s_t = i) + \pi_i[\Phi(i) \otimes \mathbf{c}(i) + \mathbf{c}(i) \otimes \Phi(i)]\\ &\quad \times E(\mathbf{z}_{t-1} | s_t = i) + \pi_i \mathbf{c}(i) \otimes \mathbf{c}(i) + \pi_i E(\Sigma_t \otimes \Sigma_t | s_t = i)\\ &= \textstyle\sum_{j=1}^M p_{ji}[\Phi(i) \otimes \Phi(i)]\pi_j E(\mathbf{z}_{t-1} \otimes \mathbf{z}_{t-1} | s_{t-1} = j)\\ &\quad + \textstyle\sum_{j=1}^M p_{ji}[\Phi(i) \otimes \mathbf{c}(i) + \mathbf{c}(i) \otimes \Phi(i)]\pi_j E(\mathbf{z}_{t-1} | s_{t-1} = j)\\ &\quad + \pi_i \mathbf{c}(i) \otimes \mathbf{c}(i) + \pi_i E(\Sigma_t \otimes \Sigma_t | s_t = i).\end{aligned}$$

For $i = 1, \ldots, M$, let $\mathbf{W}_{it}$, $\mathbf{C}_i$ and $\mathbf{S}_{it}$ be the i-th block of $\mathbf{W}_t$, $\mathbf{C}$ and $\mathbf{S}_t$, respectively. For $i, j = 1, \ldots, M$, let $\mathbf{P}(\Phi \otimes \Phi)_{ij}$ and $\mathbf{D}_{ij}$ be the $(i,\ j)$ block of $\mathbf{P}(\Phi \otimes \Phi)$ and $\mathbf{D}$, respectively. Then $\mathbf{W}_{it}$, $\mathbf{C}_i$, and $\mathbf{S}_{it}$ are $n^2 \times 1$. Furthermore, $\mathbf{P}(\Phi \otimes \Phi)_{ij}$ is $n^2 \times n^2$, and $\mathbf{D}_{ij}$ is $n^2 \times n$. Using these notations, the last formula becomes

$$\mathbf{W}_{it} = \sum_{j=1}^{M} \mathbf{P}(\Phi \otimes \Phi)_{ij} \mathbf{W}_{jt} + \sum_{j=1}^{M} \mathbf{D}_{ij} \mathbf{U}_j + \mathbf{C}_i + \mathbf{S}_{it}$$

for $i = 1, \ldots, M$. This proves the claimed matrix relation. Collecting the above formulas yields

$$\left[\mathbf{I}_{Mn^2} - \mathbf{P}(\Phi \otimes \Phi)\right] \mathbf{W}_t = \mathbf{DU} + \mathbf{C} + \mathbf{S}_t,$$

hence

$$\left[\mathbf{I}_{Mn^2} - \mathbf{P}(\Phi \otimes \Phi)\right] \left(\mathbf{I}_M \otimes \tilde{\mathbf{f}}\right) \mathbf{V}_t = \mathbf{DU} + \mathbf{C} + (\mathbf{I}_M \otimes \tilde{\mathbf{g}}) \mathbf{V}_t$$

or, equivalently,

$$\left(\mathbf{I}_M \otimes \tilde{\mathbf{f}}\right) \mathbf{V}_t = \left[\mathbf{I}_{Mn^2} - \mathbf{P}(\Phi \otimes \Phi)\right]^{-1} (\mathbf{DU} + \mathbf{C}) + \left[\mathbf{I}_{Mn^2} - \mathbf{P}(\Phi \otimes \Phi)\right]^{-1} (\mathbf{I}_M \otimes \tilde{\mathbf{g}}) \mathbf{V}_t$$

by using Assumption 6. By definition of $\mathbf{Q}$, the last relation becomes

$$\mathbf{QV}_t = \left[\mathbf{I}_{Mn^2} - \mathbf{P}(\Phi \otimes \Phi)\right]^{-1} (\mathbf{DU} + \mathbf{C}).$$

If $\mathbf{Q}$ has full rank MK^2, then the $(MK^2) \times (MK^2)$ matrix $\mathbf{Q}'\mathbf{Q}$ is invertible. So $\mathbf{V}_t$ can be expressed in closed form as

$$\mathbf{V}_t = \left(\mathbf{Q}'\mathbf{Q}\right)^{-1} \mathbf{Q}' \left[\mathbf{I}_{Mn^2} - \mathbf{P}(\Phi \otimes \Phi)\right]^{-1} (\mathbf{DU} + \mathbf{C}),$$

hence $\mathbf{V}_t$ is time-invariant, and we set $\mathbf{V} = \mathbf{V}_t$ for every $t \in \mathbb{Z}$. Finally, we have $\mathrm{vec}\Sigma_{\mathbf{h}} = (\mathbf{e}' \otimes \mathbf{I}_{K^2})\mathbf{V}$, and the result follows from Theorem 2.

References

Bauwens, L., C. M. Hafner, and J. V. K. Rombouts. 2007. Multivariate Mixed Normal Conditional Heteroskedasticity. *Computational Statistics and Data Analysis* 51: 3551–3566.

Bauwens, L., A. Preminger, and J. V. K. Rombouts. 2010. Theory and Inference for a Markov Switching GARCH Model. *Econometrics Journal* 13: 218–244.

Bollerslev, T. 1986. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31: 307–327.

Cavicchioli, M. 2017. Higher Order Moments of Markov Switching VARMA Models. *Econometric Theory* 33: 1502–1515.

Comte, F., and O. Lieberman. 2003. Asymptotic Theory for Multivariate GARCH Processes. *Journal of Multivariate Analysis* 84: 61–84.

Engle, R. F. 1982. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of the United Kingdom Inflation. *Econometrica* 50: 987–1007.

Engle, R. F., and K. F. Kroner. 1995. Multivariate Simultaneous Generalized ARCH. *Econometric Theory* 11: 122–150.

Francq, C., M. Roussignol, and J. M. Zakoïan. 2001. Conditional Heteroskedasticity Driven by Hidden Markov Chains. *Journal of Time Series Analysis* 22: 197–220.

Francq, C., and J. M. Zakoïan. 2001. Stationarity of Multivariate Markov Switching ARMA Models. *Journal of Econometrics* 102: 339–364.

Francq, C., and J. M. Zakoïan. 2005. The L^2-Structures of Standard and Switching-Regime GARCH Models. *Stochastic Processes and Their Applications* 115: 1557–1582.

Francq, C., and J. M. Zakoïan. 2011. *GARCH Models: Structure, Statistical Inference and Financial Applications*. Chichester: John Wiley & Sons.

Francq, C., and J. M. Zakoïan. 2012. QML Estimation of a Class of Multivariate Asymmetric GARCH Models. *Econometric Theory* 28: 179–206.

Haas, M., S. Mittnik, and M. S. Paolella. 2004. Mixed Normal Conditional Heteroskedasticity. *Journal of Financial Econometrics* 2: 211–250.

Hafner, C. M. 2003. Fourth Moment Structure of Multivariate GARCH Models. *Journal of Financial Econometrics* 1: 26–54.

Hafner, C. M., and A. Preminger. 2009. On Asymptotic Theory of Multivariate GARCH Models. *Journal of Multivariate Analysis* 100: 2044–2054.

Hamilton, J. D. 1994. *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Jones, G. L. 2004. On the Markov Chain Central Limit Theorem. *Probability Surveys* 1: 299–320.

Lee, H.-T., and J. K. Yoder. 2007. A Bivariate Markov Regime Switching GARCH Approach to Estimate Time Varying Minimum Variance Hedge Ratios. *Applied Economics* 39: 1253–1265.

Liu, J. C. 2006. Stationarity of a Markov-Switching GARCH Model. *Journal of Financial Econometrics* 4: 573–593.

Lütkepohl, H. 2007. *New Introduction to Multiple Time Series Analysis*, 2nd edn. New York: Springer-Verlag Berlin-Heidelberg.

Magnus, J. R. 1988. *Linear Structures*. London: Griffin.

Magnus, J. R., and H. Neudecker. 1986. Symmetry, 0–1 Matrices and Jacobians. A Review. *Econometric Theory* 2: 157–190.

Mardia, K. V. 1970. Measures of Multivariate Skewness and Kurtosis with Applications. *Biometrika* 57: 519–530.

Nakatani, T., and T. Teräsvirta. 2008. Positivity Constraints on the Conditional Variances in the Family of Conditional Correlation GARCH Models. *Finance Research Letters* 5: 88–95.

Zhang, Z., W. K. Li, and K. C. Yuen. 2006. On a Mixture GARCH Time-Series Model. *Journal of Time Series Analysis* 27: 577–597.