Haas, Markus; Liu, Ji-Chun

Conference Paper

Theory for a Multivariate Markov--switching GARCH Model with an Application to Stock Markets

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Financial Econometrics, No. B22-V2

Provided in Cooperation with:
Verein für Socialpolitik / German Economic Association

Suggested Citation: Haas, Markus; Liu, Ji-Chun (2015) : Theory for a Multivariate Markov--switching GARCH Model with an Application to Stock Markets, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Financial Econometrics, No. B22-V2, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft

This Version is available at:
http://hdl.handle.net/10419/112855

Theory for a Multivariate Markov–switching GARCH Model with an Application to Stock Markets

Abstract

We consider a multivariate Markov–switching GARCH model which allows for regime–specific volatility dynamics, leverage effects, and correlation structures. Stationarity conditions are derived, and consistency of the maximum likelihood estimator (MLE) is established under the assumption of Gaussian innovations. A Lagrange Multiplier (LM) test for correct specification of the correlation dynamics is devised, and a simple recursion for computing multi–step–ahead conditional covariance matrices is provided. The theory is illustrated with an application to global stock market and real estate equity returns. The empirical analysis highlights the importance of the conditional distribution in Markov–switching time series models. Specifications with Student's t innovations dominate their Gaussian counterparts both in– and out–of–sample. The dominating specification appears to be a two–regime Student's t process with correlations which are higher in the turbulent (high–volatility) regime.

JEL classification: C32; C51; C58

Keywords—conditional volatility, covariance forecasts, Markov–switching, multivariate GARCH

1 Introduction

Asset return distributions are typically characterized by fat tails, conditional heteroskedasticity, and nonlinear dependence. Regarding the latter, a frequent concern is that the dependence between assets increases in periods of market turbulence. This has serious implications for portfolio and risk management, because it means that "the benefits of diversification are partly eroded when they are needed most" (Campbell et al., 2002). An overview over the extensive literature studying this phenomenon and further evidence is provided, e.g., by Kasch and Caporin (2013) and Mittnik (2014). For example, Kasch and Caporin (2013) develop a multivariate GARCH model with dynamic correlations being allowed to depend on conditional volatility and, for major stock markets, find that "turbulent periods coincide with an increase in cross–market comovement."

Markov–switching models (MSMs) are able to capture all of the aforementioned *stylized facts* of asset return distributions, and their use is very popular in financial modeling because, in addition to their flexibility, "the idea of regime changes is natural and intuitive" (Ang and Timmermann, 2012). For example, in bearish markets, expected returns, conditional volatilities and their dynamics, and correlations can differ from their respective counterparts in more normal or bullish market periods. Regime–specific dynamics may also be related to various types of trading patterns, as represented by "information" and "feedback" traders (Dean and Faff, 2008). See Guidolin (2011) and Ang and Timmermann (2012) for an overview over the many applications of MSMs.

In this paper, we investigate the properties of a multivariate extension of the Markov–switching (MS) GARCH model of Haas et al. (2004), allowing for regime–specific volatility dynamics, leverage effects, and correlation structures. Stationarity conditions are derived, and consistency of the maximum likelihood estimator (MLE) is established under the assumption of Gaussian innovations. The latter result has been unknown so far even for the univariate version of this model and is thus of more general interest. The model we consider assumes constant conditional within–regime correlations. Among other convenient features, this property allows us to derive a simple recursion for multi–step–ahead conditional covariance matrices for mean–variance portfolio allocation. However, one will want to test whether this assumption is also justified empirically. Thus a test against within–regime correlation dynamics is proposed, adopting Hamilton's (1996) Lagrange Multiplier (LM) framework along with Tse's (2000) test for constant conditional correlations in a single–regime GARCH model. The theory is illustrated with an application to global stock market and real estate equity returns.

The empirical analysis highlights the importance of the conditional distribution in MS time series models. Namely, since the conditional distribution in MSMs with Gaussian regimes is a discrete mixture of normals and thus already thick–tailed, one might guess that use of a more flexible within–regime distribution is unnecessary in this framework. Indeed, as observed by Guidolin (2011), "it seems that most authors are still finding that traditional Gaussian mixture models are generally sufficient to the task assigned to MSMs." However, in our application, specifications with Student's t innovations dominate their Gaussian counterparts both in– and out–of–sample. In particular, as discussed in Section 4, the Gaussian specification turns out to suffer from its inability to correctly track the regime–switching process. The dominating specification appears to be a two–regime Student's t process with correlations which are higher in the turbulent (high–volatility) regime.

The structure of the paper is as follows. In Section 2, we define the model and discuss its relation to the literature. The statistical properties are presented in Section 3. Section 4 provides an application to financial data, and Section 5 concludes. Proofs of theorems and technical details of the LM test against misspecification of conditional correlations are gathered in Appendices A and B, respectively.

2 Definition of the process

The multivariate Markov–switching (MS) GARCH process introduced in this section generalizes the univariate model proposed in Haas et al. (2004). For alternative approaches to MS GARCH processes, see, e.g., Gray (1996), Dueker (1997), Klaassen (2002), and Augustyniak (2014), as well as the review in Haas and Paolella (2012). Liu (2007) extended the model of Haas et al. (2004) to allow for an asymmetric response of volatility to positive and negative shocks, which is also incorporated in the model discussed herein.

Let the M–dimensional time series $\{\boldsymbol{\epsilon}_t\}$ satisfy

$$\boldsymbol{\epsilon}_t = \boldsymbol{D}_{\Delta_t,t} \cdot \boldsymbol{z}_t, \tag{1}$$

where $\{\Delta_t\}$ is a Markov chain with finite state space $\mathscr{E} = \{1,\dots,k\}$ and irreducible and aperiodic transition matrix $\boldsymbol{P}$,

$$\boldsymbol{P} = \begin{pmatrix} p_{11} & \cdots & p_{k1} \\ \vdots & \cdots & \vdots \\ p_{k1} & \cdots & p_{kk} \end{pmatrix}, \tag{2}$$

where the transition probabilities $p_{ij} = p(\Delta_t = j|\Delta_{t-1} = i)$, $i, j \in \mathscr{E}$, and the stationary distribution of the chain is denoted by $\boldsymbol{\pi}_\infty = (\pi_{1,\infty}, \pi_{2,\infty}, \ldots, \pi_{k,\infty})'$. Matrix $\boldsymbol{D}_{\Delta_t,t} = \mathrm{diag}(\boldsymbol{\sigma}_{\Delta_t,t})$, where $\boldsymbol{\sigma}_{jt} = (\sigma_{1jt}, \ldots, \sigma_{Mjt})' \in \mathbb{R}^M$, $j \in \mathscr{E}$, contains the regime–specific conditional standard deviations of the elements of $\boldsymbol{\epsilon}_t$. Moreover,

$$\boldsymbol{z}_t = \boldsymbol{R}_{\Delta_t}^{1/2} \boldsymbol{\xi}_t, \tag{3}$$

where $\boldsymbol{R}_j = (\rho_{\ell m,j})_{\ell,m=1,\ldots,M}$, $j = 1, \ldots, k$, is a (regime–specific) correlation matrix, and $\{\boldsymbol{\xi_t}\}$ is a sequence of iid random vectors with zero mean and identity covariance matrix. In the applications below, we assume that $\boldsymbol{\xi_t}$ has a Student's t distribution with $\nu > 2$ degrees of freedom, with density given by (B.5) in Appendix B, i.e.,

$$\boldsymbol{\xi_t} \overset{iid}{\sim} t(\mathbf{0}, \boldsymbol{I}_M, \nu), \tag{4}$$

which includes normality as a limiting case ($\nu \to \infty$). $\{\Delta_t\}$ and $\{\boldsymbol{\xi_t}\}$ are assumed to be independent.

The regime–specific conditional standard deviations follow simultaneous asymmetric absolute value GARCH(1,1) (AGARCH) processes, i.e., in the most general form,

$$\begin{aligned} \boldsymbol{\sigma}_{jt} &= \boldsymbol{\omega}_j + \boldsymbol{A}_j|\boldsymbol{\epsilon}_{t-1}| - (\boldsymbol{A}_j \odot \boldsymbol{\Gamma}_j)\boldsymbol{\epsilon}_{t-1} + \boldsymbol{B}_j\boldsymbol{\sigma}_{j,t-1} & (5)\\ &= \boldsymbol{\omega}_j + (\boldsymbol{A}_j|\boldsymbol{Z}_t| - (\boldsymbol{A}_j \odot \boldsymbol{\Gamma}_j)\boldsymbol{Z}_t)\boldsymbol{\sigma}_{\Delta_{t-1},t-1} + \boldsymbol{B}_j\boldsymbol{\sigma}_{j,t-1} \quad j \in \mathscr{E}, & (6) \end{aligned}$$

where $\boldsymbol{Z}_t = \mathrm{diag}(\boldsymbol{z}_t)$, a matrix in absolute value bars means that the absolute value of each element is taken, $\boldsymbol{\omega}_j = (\omega_{1j}, \ldots, \omega_{Mj})'$, and

$$\boldsymbol{A}_j = [a_{\ell m,j}]_{\ell,m=1,\ldots,M}, \quad \boldsymbol{\Gamma}_j = [\gamma_{\ell m,j}]_{\ell,m=1,\ldots,M}, \quad \boldsymbol{B}_j = [b_{\ell m,j}]_{\ell,m=1,\ldots,M}, \quad j \in \mathscr{E}. \tag{7}$$

Parameters $\gamma_{\ell m,j} \in (-1, 1)$, $\ell, m = 1, \ldots, M$, $j \in \mathscr{E}$, allow the conditional standard deviations to react asymmetrically to positive and negative news of the same magnitude as in Ding et al. (1993).

The model defined by (1)–(7), which will be referred to as a k–component Markov–switching Constant Conditional Correlation GARCH process, or, in short, MS(k) CCC–GARCH, is an asymmetric multi–regime version of the extended CCC (ECCC) model studied by Jeantheau (1998), which itself generalizes the CCC of Bollerslev (1990) by allowing for volatility interactions, which are often of interest in finance and macroeconomics (e.g., Nakatani and Teräsvirta, 2009; and Conrad and Karanasos, 2010). In most applications the diagonal model, with all $\boldsymbol{A}_j$, $\boldsymbol{B}_j$, and $\boldsymbol{\Gamma}_j$ being diagonal matrices, will be preferred for reasons of parsimony; an ARCH version of such a model was used by Ramchand and Susmel (1998). Extensions of the ECCC to allow for (possibly cross) leverage effects along the lines of Glosten et al. (1993) are explored in McAleer et al. (2009) and Francq and Zakoïan (2012).

The specification of the volatility dynamics (5)–(6) in terms of the conditional standard deviation instead of the conditional variance, as originally proposed by Taylor (1986), serves two purposes: First, empirically, it appears that it typically improves the fit as compared to the formulation in terms of the conditional variance, and is very often close to the MLE when the "power parameter" (as in Ding et al., 1993) ist freely estimated from the data (e.g., Giot and Laurent, 2003; Lejeune, 2009; and Broda et al., 2013). Second, as noted by Pelletier (2006), this specification allows for closed-form calculation of multi–step–ahead conditional covariance matrices, as required, e.g., for mean–variance portfolio optimization over horizons longer than one period. The model suggested by Pelletier (2006), referred to as the regime–switching dynamic correlation (RSDC) model, is nested in (1)–(7) when only the conditional correlation matrices are subject to regime–switching, i.e., in (5)–(6), $\boldsymbol{\omega}_1 = \cdots = \boldsymbol{\omega}_k$, $\boldsymbol{A}_{\mathbf{1}} = \cdots = \boldsymbol{A}_k$, and $\boldsymbol{B}_1 = \cdots = \boldsymbol{B}_k$. Covariance matrix forecasts for this restricted model are considered in Pelletier (2006) und Haas (2010), whereas a convenient scheme for forecasting the general model (1)–(7) will be developed in Section 3.3.

In (5)–(6), conditions have to be imposed to make sure that all elements of $\boldsymbol{\sigma}_{jt}$ remain positive with probability 1, $j = 1, \ldots, k$. As observed by He and Teräsvirta (2004), an obvious sufficient condition is that $\boldsymbol{\omega}_j > 0$ and $\boldsymbol{A}_j, \boldsymbol{B}_j \geq 0$ elementwise, but this is not necessary with nondiagonal $\boldsymbol{A}_j$ (Nakatani and Teräsvirta, 2008; Conrad and Karanasos, 2010). For the diagonal model, which is of particular importance in the applications, $\boldsymbol{\omega}_j > 0$ and $\boldsymbol{A}_j, \boldsymbol{B}_j \geq 0$, $j = 1, \ldots, k$, are necessary, however. In the discussion of the MLE in Section 3.4, we will also introduce an identifiability condition in Assumption 1.

Regarding the distribution of the innovations $\{\boldsymbol{\xi}_t\}$, note that (4) includes Gaussian innovations as a limiting case, when $\nu \to \infty$. Though normality is still the dominant distributional assumption in regime–switching models (cf. Guidolin, 2011), allowing for fat–tailed innovations can improve both in–sample fit and out–of–sample forecasting performance of MS GARCH models, as pointed out, e.g., by Klaassen (2002), and Ardia (2009); see Section 4 for a detailed discussion and illustration.[1]

[1] Time series models of independently switching Student's t components for financial data are considered, e.g., in Giacomini et al. (2008), Wong et al. (2009), and Nikolaev et al. (2013).

3 Properties of the model

In this section, we investigate the statistical properties of the MS(k) CCC–GARCH process. In particular, we present conditions for strict stationarity and the existence of unconditional moments in Section 3.1. Explicit formulas for moments of frequent interest are provided in 3.2, namely the unconditional covariance matrix and the autocorrelations of the absolute values, which can be used to characterize the joint volatility dynamics. Moreover, a simple recursive scheme for obtaining multi–step–ahead covariance matrices is derived in Section 3.3, fostering applications to mean–variance portfolio selection in environments with changing volatilities and correlations. Finally, in Section 3.4, we discuss the maximum likelihood estimator and establish its consistency for the case of normally distributed $\{\boldsymbol{\xi}_t\}$ in (3).

To set out the properties of the MS(k) CCC–GARCH process, we define the matrices

$$\boldsymbol{X}_t = \begin{pmatrix} \boldsymbol{\sigma}_{1t} \\ \vdots \\ \boldsymbol{\sigma}_{kt} \end{pmatrix}, \quad \boldsymbol{\omega} = \begin{pmatrix} \boldsymbol{\omega}_1 \\ \vdots \\ \boldsymbol{\omega}_k \end{pmatrix}, \quad \boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_1 \\ \vdots \\ \boldsymbol{A}_k \end{pmatrix}, \quad \tilde{\boldsymbol{A}} = \begin{pmatrix} \boldsymbol{A}_1 \odot \boldsymbol{\Gamma}_1 \\ \vdots \\ \boldsymbol{A}_k \odot \boldsymbol{\Gamma}_k \end{pmatrix},$$

and $\boldsymbol{B} = \text{blockdiag}(\boldsymbol{B}_1, \ldots, \boldsymbol{B}_k) = \bigoplus_{j=1}^k \boldsymbol{B}_j$. This gives rise to the representation

$$\boldsymbol{X}_t = \boldsymbol{\omega} + \boldsymbol{C}_{\Delta_{t-1},t-1}\boldsymbol{X}_{t-1}, \tag{8}$$

where

$$\boldsymbol{C}_{\Delta_t,t} = (\boldsymbol{A}|\boldsymbol{Z}_t| - \tilde{\boldsymbol{A}}\boldsymbol{Z}_t)(\boldsymbol{e}'_{\Delta_t} \otimes \boldsymbol{I}_M) + \boldsymbol{B}, \tag{9}$$

and $\boldsymbol{e}_j$ is the jth unit vector in $\mathbb{R}^k$, $j = 1, \ldots, k$.

3.1 Stationarity and existence of moments

We first provide a necessary and sufficient condition for the existence of a strictly stationary solution of the MS(k) CCC–GARCH process. Theorem 1 generalizes results for the associated univariate MS GARCH process in Liu (2006, 2007).[2]

Theorem 1. *The MS(k) CCC-GARCH(1,1) process defined by (1)–(7) has a unique strictly stationary and ergodic solution if and only if the top Lyapunov exponent γ_C associated to the*

[2] Francq et al. (2001) and Francq and Zakoïan (2005, 2008) consider stationarity and moment properties of an alternative univariate MS GARCH process; see also Abramson and Cohen (2007).

random matrices ($\boldsymbol{C}_{\Delta_t,t}$) *is strictly negative. Moreover, this stationary solution is explicitly expressed as*

$$\boldsymbol{\epsilon}_t = \left[\operatorname{diag}\left((\boldsymbol{e}'_{\Delta_t} \otimes \boldsymbol{I}_M)\left(\boldsymbol{\omega} + \sum_{n=1}^{\infty} \boldsymbol{C}_{\Delta_{t-1},t-1}\boldsymbol{C}_{\Delta_{t-2},t-2}\ldots\boldsymbol{C}_{\Delta_{t-n+1},t-n+1}\boldsymbol{\omega}\right)\right)\right]^{1/2} \boldsymbol{R}^{1/2}_{\Delta_t}\boldsymbol{\xi}_t.$$

The condition in Theorem 1 may be inconvenient to check in practice. Theorem 2 offers an alternative criterion which is easier to handle and provides additional information about the moment structure of the process. To state this criterion, we define the matrices

$$\begin{aligned} \boldsymbol{C}_1(j) &= \mathrm{E}(\boldsymbol{C}_{jt}|\Delta_t = j), \quad \boldsymbol{C}_2(j) = \mathrm{E}(\boldsymbol{C}_{jt} \otimes C_{jt}|\Delta_t = j), \quad \ldots, \\ \boldsymbol{C}_l(j) &= \mathrm{E}(\boldsymbol{C}_{jt}^{\otimes l}|\Delta_t = j), \quad j \in \mathscr{E}, \quad l \in \mathbb{N}. \end{aligned} \tag{10}$$

Furthermore, we adopt the following notation from Francq and Zakoïan (2005): For any function $f : \mathscr{E} \mapsto M_{n\times n'}(\mathbb{R})$, where $M_{n\times n'}(\mathbb{R})$ is the space of real $n \times n'$ matrices, and $\mathscr{E} = \{1, \ldots, k\}$ is the state space of $\{\Delta_t\}$, define the matrix

$$\mathbb{P}_f = \begin{pmatrix} p_{11}f(1) & \cdots & p_{k1}f(1) \\ \vdots & \cdots & \vdots \\ p_{1k}f(k) & \cdots & p_{kk}f(k) \end{pmatrix}. \tag{11}$$

Theorem 2. *Suppose that the l-th moments of* ($\boldsymbol{\xi}_t$) *are finite and*

$$\varrho(\mathbb{P}_{C_l}) < 1,$$

where $\varrho(\mathbb{P}_{C_l})$ denotes the spectral radius of $\mathbb{P}_{C_l}$ defined in (11), and l is a strictly positive integer. Then (1)–(7) has a unique strictly stationary and ergodic solution ($\boldsymbol{\epsilon}_t$)*, and the l-th absolute moments of* ($\boldsymbol{\epsilon}_t$) *are finite.*

For example, the matrices required by Theorem 2 to check for the first moment are given by

$$\boldsymbol{C}_1(j) = \kappa_1\boldsymbol{A}(\boldsymbol{e}'_j \otimes \boldsymbol{I}_M) + \boldsymbol{B},$$

where

$$\kappa_1 = \mathrm{E}(|z_{it}|) = \begin{cases} \sqrt{\frac{2}{\pi}} & \text{if } z_{it} \sim \mathrm{N}(0,1) \\ \frac{\sqrt{\nu-2}\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{\pi}\Gamma(\nu/2)} & \text{if } z_{it} \sim t_\nu(0,1). \end{cases}$$

To check the condition for covariance stationarity, we need the (regime–specific) second moment matrices of the absolute innovations, i.e.,

$$\tilde{\boldsymbol{R}}_j := \mathrm{E}(|\boldsymbol{z}_t\boldsymbol{z}'_t||\Delta_t = j), \quad j \in \mathscr{E},$$

the elements of which are provided by the result of Nabeya (1951) that, for bivariate standard normal x and y with correlation ρ, we have

$$\mathrm{E}(|xy|) = \frac{2}{\pi}(\sqrt{1-\rho^2} + \rho \arcsin \rho). \tag{12}$$

Equation (12) continues to hold for a unit–variance bivariate Student's t distribution, as detailed in the Appendix of Haas (2010). Moreover, let

$$\begin{aligned} \boldsymbol{\Omega}(j) &= \mathrm{E}(\boldsymbol{Z}_t \otimes \boldsymbol{Z}_t | \Delta_t = j) = \mathrm{diag}(\mathrm{vec}(\boldsymbol{R}_j)), \\ \tilde{\boldsymbol{\Omega}}(j) &= \mathrm{E}(|\boldsymbol{Z}_t| \otimes |\boldsymbol{Z}_t| | \Delta_t = j) = \mathrm{diag}(\mathrm{vec}(\tilde{\boldsymbol{R}}_j)), \quad j \in \mathscr{E}. \end{aligned}$$

Then matrices $\boldsymbol{C}_2(j)$, $j \in \mathscr{E}$, are given by

$$\begin{aligned} \boldsymbol{C}_2(j) &= \Big((\boldsymbol{A} \otimes \boldsymbol{A})\tilde{\boldsymbol{\Omega}}(j) + (\tilde{\boldsymbol{A}} \otimes \tilde{\boldsymbol{A}})\boldsymbol{\Omega}(j) \Big) \left(\boldsymbol{e}_j' \otimes \boldsymbol{I}_M \otimes \boldsymbol{e}_j' \otimes \boldsymbol{I}_M \right) \\ &\quad + \kappa_1 \left(\boldsymbol{e}_j' \otimes \boldsymbol{A} \otimes \boldsymbol{B} + \boldsymbol{B} \otimes \boldsymbol{e}_j' \otimes \boldsymbol{A} \right) + \boldsymbol{B} \otimes \boldsymbol{B}. \end{aligned}$$

For later reference, we also define, for $j \in \mathscr{E}$,

$$\Upsilon(j) = \mathrm{E}\left(\boldsymbol{I}_k \otimes \boldsymbol{Z}_t \otimes \boldsymbol{I}_k \otimes \boldsymbol{Z}_t | \Delta_t = j \right) = \mathrm{diag}[\mathrm{vec}((\mathbf{1}_k \mathbf{1}_k') \otimes \boldsymbol{R}_j)], \tag{13}$$

$$\tilde{\Upsilon}(j) = \mathrm{E}\left(\boldsymbol{I}_k \otimes |\boldsymbol{Z}_t| \otimes \boldsymbol{I}_k \otimes |Z_t| | \Delta_t = j \right) = \mathrm{diag}[\mathrm{vec}((\mathbf{1}_k \mathbf{1}_k') \otimes \tilde{\boldsymbol{R}}_j)], \tag{14}$$

where $\mathbf{1}_k$ is a k–dimensional column of ones.

3.2 Calculation of the moments

It may be of interest to calculate the moments of a specific MS CCC–GARCH process. In particular, we are interested in the overall and the regime–specific unconditional covariance and correlation matrices as well as in the dynamic correlation structure of the absolute values of the process. We use the following basic result.

Lemma 1. *(Francq and Zakoïan, 2005, Lemma 3) For $\ell \geq 1$, if the variable $\boldsymbol{Y}_{t-\ell}$ belongs to the information set generated by $\{\epsilon_s : s \leq t - \ell\}$, then*

$$\pi_{j,\infty} \mathrm{E}(\boldsymbol{Y}_{t-\ell} | \Delta_t = j) = \sum_{i=1}^{k} \pi_{i,\infty} p_{ij}^{(\ell)} \mathrm{E}(\boldsymbol{Y}_{t-\ell} | \Delta_{t-\ell} = i),$$

where the $p_{ij}^{(\ell)} := p(\Delta_t = j | \Delta_{t-\ell} = i)$, $i, j \in \mathscr{E}$, denote the ℓ–step transition probabilities, as given by the elements of $\boldsymbol{P}^\ell$.

Using Lemma 1, we have

$$\begin{aligned}\pi_{j,\infty}\,\mathrm{E}(\boldsymbol{X}_t|\Delta_{t-1}=j) &= \pi_{j,\infty}\boldsymbol{\omega}+\pi_{j,\infty}\boldsymbol{C}_1(j)\,\mathrm{E}(\boldsymbol{X}_{t-1}|\Delta_{t-1}=j) \qquad (15)\\ &= \pi_{j,\infty}\boldsymbol{\omega}+\sum_{i=1}^{k}p_{ij}\boldsymbol{C}_1(j)\pi_{i,\infty}\,\mathrm{E}(\boldsymbol{X}_{t-1}|\Delta_{t-2}=i),\quad j=1,\dots,k.\end{aligned}$$

Equation (15) implies

$$\boldsymbol{V}_1=\boldsymbol{\pi}_\infty\otimes\boldsymbol{\omega}+\mathbb{P}_{C_1}\boldsymbol{V}_1,$$

where

$$\boldsymbol{V}_1=\begin{pmatrix}\pi_{1,\infty}\,\mathrm{E}(\boldsymbol{X}_t|\Delta_{t-1}=1)\\ \pi_{2,\infty}\,\mathrm{E}(\boldsymbol{X}_t|\Delta_{t-1}=2)\\ \vdots\\ \pi_{k,\infty}\,\mathrm{E}(\boldsymbol{X}_t|\Delta_{t-1}=k)\end{pmatrix}. \qquad (16)$$

Thus the first absolute moments are

$$\begin{aligned}\mathrm{E}(|\boldsymbol{\epsilon}_t|) &= \sum_{j=1}^{k}\pi_{j,\infty}\,\mathrm{E}(|\boldsymbol{\epsilon}_t||\Delta_t=j)\\ &= \kappa_1\sum_{j=1}^{k}\sum_{i=1}^{k}\pi_{j,\infty}p(\Delta_{t-1}=i|\Delta_t=j)\,\mathrm{E}(\boldsymbol{\sigma}_{jt}|\Delta_{t-1}=i)\\ &= \kappa_1\sum_{j=1}^{k}\sum_{i=1}^{k}p_{ij}\pi_{i,\infty}\,\mathrm{E}(\boldsymbol{\sigma}_{jt}|\Delta_{t-1}=i)\\ &= \kappa_1(\mathrm{vec}(\boldsymbol{P})'\otimes\boldsymbol{I}_M)\boldsymbol{V}_1.\end{aligned}$$

For the covariance matrix, proceeding similarly,

$$\begin{aligned}\pi_{j,\infty}\,\mathrm{E}[\mathrm{vec}(\boldsymbol{X}_t\boldsymbol{X}_t')|\Delta_{t-1}=j] &= \pi_{j,\infty}(\boldsymbol{\omega}\otimes\boldsymbol{\omega})+\sum_{i=1}^{k}p_{ij}\boldsymbol{C}_{21}(j)\,\mathrm{E}(\boldsymbol{X}_{t-1}|\Delta_{t-2}=i)\\ &\quad+\sum_{i=1}^{k}p_{ij}\boldsymbol{C}_2(j)\,\mathrm{E}[\mathrm{vec}(\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}')|\Delta_{t-2}=i], \qquad (17)\end{aligned}$$

where $\boldsymbol{C}_{21}(j)=\boldsymbol{\omega}\otimes\boldsymbol{C}_1(j)+\boldsymbol{C}_1(j)\otimes\boldsymbol{\omega}$, $j=1,\dots,k$. Equation (17) implies

$$\boldsymbol{V}_2=\boldsymbol{\pi}_\infty\otimes\boldsymbol{\omega}\otimes\boldsymbol{\omega}+\mathbb{P}_{C_{21}}\boldsymbol{V}_1+\mathbb{P}_{C_2}\boldsymbol{V}_2,$$

where $\boldsymbol{V}_2$ is as $\boldsymbol{V}_1$ in (16) but with $\pi_{j,\infty}\,\mathrm{E}(\boldsymbol{X}_t|\Delta_{t-1}=j)$ replaced by $\pi_{j,\infty}\,\mathrm{E}[\mathrm{vec}(\boldsymbol{X}_t\boldsymbol{X}_t')|\Delta_{t-1}=$

$j]$, $j = 1, \ldots, k$. Thus the unconditional covariance matrix of $\{\boldsymbol{\epsilon}_t\}$ is

$$\begin{aligned}
\mathrm{E}[\mathrm{vec}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t')] &= \sum_{j=1}^{k} \pi_{j,\infty}\, \mathrm{E}[\mathrm{vec}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t')|\Delta_t = j] \\
&= \sum_{j=1}^{k} \pi_{j,\infty}\, \mathrm{E}\left\{\mathrm{vec}[(\boldsymbol{e}_j \otimes \boldsymbol{I}_M)'(\boldsymbol{I}_k \otimes \boldsymbol{Z}_t)\boldsymbol{X}_t\boldsymbol{X}_t'(\boldsymbol{I}_k \otimes \boldsymbol{Z}_t)(\boldsymbol{e}_j \otimes \boldsymbol{I}_M)|\Delta_t = j]\right\} \\
&= \sum_{j=1}^{k} (\boldsymbol{e}_j \otimes \boldsymbol{I}_M \otimes \boldsymbol{e}_j \otimes \boldsymbol{I}_M)' \sum_{i=1}^{k} p_{ij}\Upsilon(j)\pi_{i,\infty}\, \mathrm{E}[\mathrm{vec}(\boldsymbol{X}_t\boldsymbol{X}_t')|\Delta_{t-1} = i] \\
&= \sum_{j=1}^{k} \boldsymbol{e}_j' \otimes (\boldsymbol{e}_j \otimes \boldsymbol{I}_M \otimes \boldsymbol{e}_j \otimes \boldsymbol{I}_M)'\mathbb{P}_{\Upsilon}\boldsymbol{V}_2, \qquad (18)
\end{aligned}$$

where definitions (11) and (13) were used. The *regime–specific* unconditional covariance matrices are also of interest and given by

$$\mathrm{E}[\mathrm{vec}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t')|\Delta_t = j] = \pi_{j,\infty}^{-1}\boldsymbol{e}_j' \otimes (\boldsymbol{e}_j \otimes \boldsymbol{I}_M \otimes \boldsymbol{e}_j \otimes \boldsymbol{I}_M)'\mathbb{P}_{\Upsilon}\boldsymbol{V}_2, \quad j = 1, \ldots, k.$$

To calculate the autocorrelation function of the absolute process, $\mathrm{E}[\mathrm{vec}(|\boldsymbol{\epsilon}_t||\boldsymbol{\epsilon}_t|')]$ is required, which directly follows from (14) and (18) as

$$\mathrm{E}[\mathrm{vec}(|\boldsymbol{\epsilon}_t||\boldsymbol{\epsilon}_t|')] = \sum_{j=1}^{k} \boldsymbol{e}_j' \otimes (\boldsymbol{e}_j \otimes \boldsymbol{I}_M \otimes \boldsymbol{e}_j \otimes \boldsymbol{I}_M)'\mathbb{P}_{\tilde{\Upsilon}}\boldsymbol{V}_2.$$

The cross moment matrices are obtained via

$$\begin{aligned}
&\mathrm{E}(|\boldsymbol{\epsilon}_t||\boldsymbol{\epsilon}_{t-\tau}|') = \mathrm{E}\left\{(\boldsymbol{e}_{\Delta_t} \otimes \boldsymbol{I}_M)'(\boldsymbol{I}_k \otimes |\boldsymbol{Z}_t|)\boldsymbol{X}_t\boldsymbol{X}_{t-\tau}'(\boldsymbol{I}_k \otimes |\boldsymbol{Z}_{t-\tau}|)(\boldsymbol{e}_{\Delta_{t-\tau}} \otimes \boldsymbol{I}_M)\right\} \\
&= \kappa_1 \sum_{i=1}^{k}\sum_{j=1}^{k} (\boldsymbol{e}_j \otimes \boldsymbol{I}_M)' p(\Delta_{t-\tau} = i \cap \Delta_t = j) \\
&\qquad \times \mathrm{E}\left\{\boldsymbol{X}_t\boldsymbol{X}_{t-\tau}'(\boldsymbol{I}_k \otimes |\boldsymbol{Z}_{t-\tau}|)|\Delta_{t-\tau} = i \cap \Delta_t = j\right\}(\boldsymbol{e}_i \otimes \boldsymbol{I}_M) \\
&= \kappa_1 \sum_{i=1}^{k}\sum_{j=1}^{k} (\boldsymbol{e}_j \otimes \boldsymbol{I}_M)'\boldsymbol{S}_{ij}(\tau)(\boldsymbol{e}_i \otimes \boldsymbol{I}_M), \qquad (19)
\end{aligned}$$

where, for $i, j = 1, \ldots, k$, and with $p(\Delta_{t-\tau} = i \cap \Delta_t = j) = \pi_{i,\infty} p_{ij}^{(\tau)}$,

$$
\begin{aligned}
\boldsymbol{S}_{ij}(\tau) &= \pi_{i,\infty} p_{ij}^{(\tau)} \operatorname{E}\left\{\boldsymbol{X}_t \boldsymbol{X}'_{t-\tau}(\boldsymbol{I}_k \otimes |\boldsymbol{Z}_{t-\tau}|) | \Delta_{t-\tau} = i \cap \Delta_t = j\right\} \\
&= \pi_{i,\infty} p_{ij}^{(\tau)} \operatorname{E}\left\{(\boldsymbol{\omega} + \boldsymbol{C}_{\Delta_{t-1},t-1}\boldsymbol{X}_{t-1}) \boldsymbol{X}'_{t-\tau}(\boldsymbol{I}_k \otimes |\boldsymbol{Z}_{t-\tau}|) | \Delta_{t-\tau} = i \cap \Delta_t = j\right\} \\
&= \pi_{i,\infty} p_{ij}^{(\tau)} \boldsymbol{\omega} \kappa_1 \operatorname{E}(\boldsymbol{X}'_{t-\tau} | \Delta_{t-\tau} = i) \\
&\quad + \pi_{i,\infty} p_{ij}^{(\tau)} \operatorname{E}\left\{\boldsymbol{C}_{\Delta_{t-1},t-1}\boldsymbol{X}_{t-1} \boldsymbol{X}'_{t-\tau}(\boldsymbol{I}_k \otimes |\boldsymbol{Z}_{t-\tau}|) | \Delta_{t-\tau} = i \cap \Delta_t = j\right\} \\
&= \pi_{i,\infty} p_{ij}^{(\tau)} \boldsymbol{\omega} \kappa_1 \operatorname{E}(\boldsymbol{X}'_{t-\tau} | \Delta_{t-\tau} = i) \\
&\quad + \sum_{\ell=1}^{k} \pi_{i,\infty} p_{i\ell}^{(\tau-1)} p_{\ell j} \\
&\qquad \times \operatorname{E}\left\{\boldsymbol{C}_{\Delta_{t-1},t-1}\boldsymbol{X}_{t-1} \boldsymbol{X}'_{t-\tau}(\boldsymbol{I}_k \otimes |\boldsymbol{Z}_{t-\tau}|) | \Delta_{t-\tau} = i \cap \Delta_{t-1} = \ell \cap \Delta_t = j\right\} \\
&= \pi_{i,\infty} p_{ij}^{(\tau)} \boldsymbol{\omega} \kappa_1 \operatorname{E}(\boldsymbol{X}'_{t-\tau} | \Delta_{t-\tau} = i) \\
&\quad + \sum_{\ell=1}^{k} \pi_{i,\infty} p_{i\ell}^{(\tau-1)} p_{\ell j} \boldsymbol{C}_1(\ell) \operatorname{E}\left\{\boldsymbol{X}_{t-1} \boldsymbol{X}'_{t-\tau}(\boldsymbol{I}_k \otimes |\boldsymbol{Z}_{t-\tau}|) | \Delta_{t-\tau} = i \cap \Delta_{t-1} = \ell\right\} \\
&= \pi_{i,\infty} p_{ij}^{(\tau)} \boldsymbol{\omega} \kappa_1 \operatorname{E}(\boldsymbol{X}'_{t-\tau} | \Delta_{t-\tau} = i) + \sum_{\ell=1}^{k} p_{\ell j} \boldsymbol{C}_1(\ell) \boldsymbol{S}_{i\ell}(\tau - 1), \quad \tau \geq 2,
\end{aligned}
$$

that is,

$$
\boldsymbol{S}(\tau) = \kappa_1 (\boldsymbol{P}^{(\tau)} \otimes \boldsymbol{\omega}) \tilde{\boldsymbol{V}}_1 + \tilde{\mathbb{P}}_{C_1} \boldsymbol{S}(\tau - 1), \quad \tau \geq 2.
$$

where

$$
\boldsymbol{S}(\tau) = \begin{pmatrix} \boldsymbol{S}_{11}(\tau) & \cdots & \boldsymbol{S}_{k1}(\tau) \\ \vdots & \ddots & \vdots \\ \boldsymbol{S}_{1k}(\tau) & \cdots & \boldsymbol{S}_{kk}(\tau) \end{pmatrix}, \quad \tilde{\boldsymbol{V}}_1 = \begin{pmatrix} \pi_{1,\infty} \operatorname{E}(\boldsymbol{X}'_t | \Delta_t = 1) & \cdots & \boldsymbol{0}_{1\times kM} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0}_{1\times kM} & \cdots & \pi_{k,\infty} \operatorname{E}(\boldsymbol{X}'_t | \Delta_t = k) \end{pmatrix},
$$

the diagonal blocks of $\tilde{\boldsymbol{V}}_1$ can be extracted from the vector $(\boldsymbol{P} \otimes \boldsymbol{I}_{kM})\boldsymbol{V}_1 = (\pi_{1,\infty} \operatorname{E}(\boldsymbol{X}_t | \Delta_t = 1)', \ldots, \pi_{k,\infty} \operatorname{E}(\boldsymbol{X}'_t | \Delta_t = k)')'$, and, similar to (11),

$$
\tilde{\mathbb{P}}_{C_1} = \begin{pmatrix} p_{11} C_1(1) & \cdots & p_{k1} C_1(k) \\ \vdots & \ddots & \vdots \\ p_{1k} C_1(1) & \cdots & p_{kk} C_1(k) \end{pmatrix}. \tag{20}
$$

For $\tau = 1$, we compute

$$
\begin{aligned}
S_{ij}(1) &= \pi_{i,\infty} p_{ij} \operatorname{E}[\boldsymbol{X}_t \boldsymbol{X}'_{t-1}(\boldsymbol{I}_k \otimes |\boldsymbol{Z}_{t-1}|) | \Delta_{t-1} = i \cap \Delta_t = j] \\
&= \pi_{i,\infty} p_{ij} \operatorname{E}[(\boldsymbol{\omega} + \boldsymbol{C}_{\Delta_{t-1},t-1}\boldsymbol{X}_{t-1}) \boldsymbol{X}'_{t-1}(\boldsymbol{I}_k \otimes |\boldsymbol{Z}_{t-1}|) | \Delta_{t-1} = i] \\
&= \kappa_1 p_{ij} \boldsymbol{\omega} \pi_{i,\infty} \operatorname{E}(\boldsymbol{X}'_t | \Delta_t = i) + p_{ij} \pi_{i,\infty} \operatorname{E}[\boldsymbol{C}_{\Delta_{t-1},t-1}\boldsymbol{X}_{t-1} \boldsymbol{X}'_{t-1}(\boldsymbol{I}_k \otimes |\boldsymbol{Z}_{t-1}|) | \Delta_{t-1} = i].
\end{aligned}
$$

Hence

$$\boldsymbol{S}(1) = \kappa_1(\boldsymbol{P} \otimes \boldsymbol{\omega})\tilde{\boldsymbol{V}}_1 + \tilde{\mathbb{P}}_{\breve{C}},$$

where $\tilde{\mathbb{P}}_{\breve{C}}$ is as in (20) with

$$\breve{\boldsymbol{C}}(i) = \pi_{i,\infty}\,\mathrm{E}[\boldsymbol{C}_{\Delta_{t-1},t-1}\boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}(\boldsymbol{I}_k \otimes |\boldsymbol{Z}_{t-1}|)|\Delta_{t-1} = i], \quad i = 1, \dots, k. \tag{21}$$

The expectation in (21) is

$$\begin{aligned}
&\mathrm{E}(\mathrm{vec}(\boldsymbol{C}_{\Delta_{t-1},t-1}\boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}(\boldsymbol{I}_M \otimes |\boldsymbol{Z}_{t-1}|))|\Delta_{t-1} = i) \\
&\quad = \pi_{i,\infty}\,\mathrm{E}((\boldsymbol{I}_M \otimes |\boldsymbol{Z}_{t-1}| \otimes \boldsymbol{C}_{\Delta_{t-1},t-1})\mathrm{vec}(\boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1})|\Delta_{t-1} = i) \\
&\quad = \mathrm{E}(\boldsymbol{I}_M \otimes |\boldsymbol{Z}_{t-1}| \otimes \boldsymbol{C}_{\Delta_{t-1},t-1})|\Delta_{t-1} = i)\pi_{i,\infty}\,\mathrm{E}(\mathrm{vec}(\boldsymbol{X}_t\boldsymbol{X}'_t)|\Delta_t = i),
\end{aligned}$$

where $\pi_{i,\infty}\,\mathrm{E}(\mathrm{vec}(\boldsymbol{X}_t\boldsymbol{X}'_t)|\Delta_t = i)$, $i = 1, \dots, k$, can be extracted from $(\boldsymbol{P} \otimes \boldsymbol{I}_{k^2M^2})\boldsymbol{V}_2 = (\pi_{1,\infty}\,\mathrm{E}(\mathrm{vec}(\boldsymbol{X}_t\boldsymbol{X}'_t)|\Delta_t = 1)', \dots, \pi_{k,\infty}\,\mathrm{E}(\mathrm{vec}(\boldsymbol{X}_t\boldsymbol{X}'_t)|\Delta_t = k)')'$, and

$$\begin{aligned}
&\mathrm{E}(\boldsymbol{I}_k \otimes |\boldsymbol{Z}_{t-1}| \otimes \boldsymbol{C}_{\Delta_{t-1},t-1})|\Delta_{t-1} = i) \\
&\quad = \mathrm{E}(\boldsymbol{I}_k \otimes |\boldsymbol{Z}_{t-1}| \otimes [(\boldsymbol{A}|\boldsymbol{Z}_{t-1}| - \tilde{\boldsymbol{A}}\boldsymbol{Z}_{t-1})(\boldsymbol{e}'_{\Delta_{t-1}} \otimes \boldsymbol{I}_M) + \boldsymbol{B}]|\Delta_{t-1} = i) \\
&\quad = \mathrm{E}(\boldsymbol{I}_k \otimes |\boldsymbol{Z}_{t-1}| \otimes (\boldsymbol{A}|\boldsymbol{Z}_{t-1}|(\boldsymbol{e}'_{\Delta_{t-1}} \otimes \boldsymbol{I}_M) + \boldsymbol{B})|\Delta_{t-1} = i) \\
&\quad = (\boldsymbol{I}_{kM} \otimes \boldsymbol{e}'_i \otimes \boldsymbol{A})\,\mathrm{E}(\boldsymbol{I}_k \otimes |\boldsymbol{Z}_{t-1}| \otimes \boldsymbol{I}_k \otimes |\boldsymbol{Z}_{t-1}||\Delta_{t-1} = i) + \kappa_1(\boldsymbol{I}_{kM} \otimes \boldsymbol{B}) \\
&\quad = (\boldsymbol{I}_{kM} \otimes \boldsymbol{e}'_i \otimes \boldsymbol{A})\tilde{\Upsilon}(i) + \kappa_1(\boldsymbol{I}_{kM} \otimes \boldsymbol{B}).
\end{aligned}$$

Finally,

$$\mathrm{E}(|\boldsymbol{\epsilon}_t||\boldsymbol{\epsilon}_{t-\tau}|') = \kappa_1(\mathrm{vec}(\boldsymbol{I}_k)' \otimes \boldsymbol{I}_M)\boldsymbol{S}(\tau)(\mathrm{vec}(\boldsymbol{I}_k) \otimes \boldsymbol{I}_M),$$

and the autocorrelation function can be computed.

3.3 Covariance matrix forecasts

Suppose we are given a current probability distribution of the chain $\boldsymbol{\pi}_t$ and an initial vector $\boldsymbol{X}_{t+1}$ (which is known at time t). Define

$$\begin{aligned}
\boldsymbol{Y}_t &= \begin{pmatrix} \boldsymbol{X}_t \\ \mathrm{vec}(\boldsymbol{X}_t\boldsymbol{X}'_t) \end{pmatrix}, \quad \tilde{\boldsymbol{\omega}} = \begin{pmatrix} \boldsymbol{\omega} \\ \boldsymbol{\omega} \otimes \boldsymbol{\omega} \end{pmatrix}, \\
\tilde{\boldsymbol{C}}_{\Delta_t,t} &= \begin{pmatrix} \boldsymbol{C}_{\Delta_t,t} & \boldsymbol{0}_{kM \times k^2M^2} \\ \boldsymbol{C}_{\Delta_t,t} \otimes \boldsymbol{\omega} + \boldsymbol{\omega} \otimes \boldsymbol{C}_{\Delta_t,t} & \boldsymbol{C}_{\Delta_t,t} \otimes \boldsymbol{C}_{\Delta_t,t} \end{pmatrix},
\end{aligned}$$

so that

$$\boldsymbol{Y}_t = \tilde{\boldsymbol{\omega}} + \tilde{\boldsymbol{C}}_{\Delta_{t-1},t-1}\boldsymbol{Y}_{t-1}. \tag{22}$$

Upon repeated substitution in (22), we can write

$$\boldsymbol{Y}_{t+d} = \sum_{\ell=1}^{d-1} \left\{ \prod_{i=1}^{\ell-1} \tilde{\boldsymbol{C}}_{\Delta_{t+d-i},t+d-i} \right\} \tilde{\boldsymbol{\omega}} + \left\{ \prod_{i=1}^{d-1} \tilde{\boldsymbol{C}}_{\Delta_{t+d-i},t+d-i} \right\} \boldsymbol{Y}_{t+1}.$$

Let $\underline{\Delta}_t = \{\Delta_s : s \leq t\}$. Then we have, taking expectations with respect to $\{\boldsymbol{\xi}_t\}$,

$$\mathrm{E}_t(\boldsymbol{Y}_{t+d}|\underline{\Delta}_{t+d-1}) = \sum_{\ell=1}^{d-1} \left\{ \prod_{i=1}^{\ell-1} \tilde{\boldsymbol{C}}(\Delta_{t+d-i}) \right\} \tilde{\boldsymbol{\omega}} + \left\{ \prod_{i=1}^{d-1} \tilde{\boldsymbol{C}}(\Delta_{t+d-i}) \right\} \boldsymbol{Y}_{t+1},$$

where, as in (10), $\tilde{\boldsymbol{C}}(j) = \mathrm{E}(\tilde{\boldsymbol{C}}_{jt}|\Delta_t = j)$. From Lemma 1 in Francq and Zakoïan (2005), we have

$$\begin{aligned}
\tilde{\boldsymbol{Y}}_t(d) \quad &:= \quad \begin{pmatrix} p_t(\Delta_{t+d-1} = 1)\, \mathrm{E}_t(\boldsymbol{Y}_{t+d}|\Delta_{t+d-1} = 1) \\ p_t(\Delta_{t+d-1} = 2)\, \mathrm{E}_t(\boldsymbol{Y}_{t+d}|\Delta_{t+d-1} = 2) \\ \vdots \\ p_t(\Delta_{t+d-1} = k)\, \mathrm{E}_t(\boldsymbol{Y}_{t+d}|\Delta_{t+d-1} = k) \end{pmatrix} \\
&= \quad \sum_{\ell=1}^{d-1} \mathbb{P}_{\tilde{C}}^{\ell-1}(\boldsymbol{\pi}_{t+d-\ell} \otimes \tilde{\boldsymbol{\omega}}) + \mathbb{P}_{\tilde{C}}^{d-1}(\boldsymbol{\pi}_t \otimes \boldsymbol{Y}_{t+1}),
\end{aligned}$$

where $\boldsymbol{\pi}_{t+d-\ell} = \boldsymbol{P}^{d-\ell}\boldsymbol{\pi}_t$. Define the matrix

$$\mathbb{I} \quad = \quad \boldsymbol{I}_k \otimes (\boldsymbol{0}_{k^2M^2 \times kM}, \boldsymbol{I}_{k^2M^2}).$$

Then the d–step–ahead covariance matrix forecast is given by

$$\begin{aligned}
\mathrm{E}_t\left(\mathrm{vec}\left(\boldsymbol{\epsilon}_{t+d}\boldsymbol{\epsilon}'_{t+d}\right)\right) \quad &= \quad \sum_{j=1}^{k} p_t(\Delta_{t+d} = j)\, \mathrm{E}_t(\mathrm{vec}(\boldsymbol{\epsilon}_{t+d}\boldsymbol{\epsilon}'_{t+d})|\Delta_{t+d} = j) \\
&= \quad \sum_{j=1}^{k} p_t(\Delta_{t+d} = j) \\
&\times \quad \mathrm{E}_t\left\{\mathrm{vec}\left[(\boldsymbol{e}_j \otimes \boldsymbol{I}_M)'(\boldsymbol{I}_k \otimes \boldsymbol{Z}_{t+d})\boldsymbol{X}_{t+d}\boldsymbol{X}'_{t+d}(\boldsymbol{I}_k \otimes \boldsymbol{Z}_{t+d})(\boldsymbol{e}_j \otimes \boldsymbol{I}_M)\right]|\Delta_{t+d} = j\right\} \\
&= \quad \sum_{j=1}^{k} p_t(\Delta_{t+d} = j)\, (\boldsymbol{e}_j \otimes \boldsymbol{I}_M \otimes \boldsymbol{e}_j \otimes \boldsymbol{I}_M)' \\
&\times \quad \mathrm{E}\left(\boldsymbol{I}_k \otimes \boldsymbol{Z}_{t+d} \otimes \boldsymbol{I}_k \otimes \boldsymbol{Z}_{t+d}|\Delta_{t+d} = j\right) \mathrm{E}_t[\mathrm{vec}(\boldsymbol{X}_{t+d}\boldsymbol{X}'_{t+d})|\Delta_{t+d} = j] \\
&= \quad \sum_{j=1}^{k} (\boldsymbol{e}_j \otimes \boldsymbol{I}_M \otimes \boldsymbol{e}_j \otimes \boldsymbol{I}_M)' \\
&\times \quad \sum_{i=1}^{k} p_{ij}\Upsilon(j) p_t(\Delta_{t+d-1} = i)\, \mathrm{E}_t[\mathrm{vec}(\boldsymbol{X}_{t+d}\boldsymbol{X}'_{t+d})|\Delta_{t+d-1} = i] \\
&= \quad \left\{ \sum_{j=1}^{k} [\boldsymbol{e}'_j \otimes (\boldsymbol{e}_j \otimes \boldsymbol{I}_M \otimes \boldsymbol{e}_j \otimes \boldsymbol{I}_M)'] \right\} \mathbb{P}_{\Upsilon}\mathbb{I}\tilde{\boldsymbol{Y}}_t(d). \qquad (23)
\end{aligned}$$

Vector $\tilde{\boldsymbol{Y}}_t(d)$ in (23) can be calculated recursively, with starting value $\tilde{\boldsymbol{Y}}_t(1) = \boldsymbol{\pi}_t \otimes \boldsymbol{Y}_{t+1}$: Namely, for $d \geq 2$,

$$\begin{aligned}
\tilde{\boldsymbol{Y}}_t(d) &= \sum_{\ell=1}^{d-1} \mathbb{P}_{\tilde{C}}^{\ell-1}(\boldsymbol{\pi}_{t+d-\ell} \otimes \tilde{\boldsymbol{\omega}}) + \mathbb{P}_{\tilde{C}}^{d-1}(\boldsymbol{\pi}_t \otimes \boldsymbol{Y}_{t+1}) \\
&= \boldsymbol{\pi}_{t+d-1} \otimes \tilde{\boldsymbol{\omega}} + \mathbb{P}_{\tilde{C}} \left\{ \sum_{\ell=1}^{d-2} \mathbb{P}_{\tilde{C}}^{\ell-1}(\boldsymbol{\pi}_{t+(d-1)-\ell} \otimes \tilde{\boldsymbol{\omega}}) + \mathbb{P}_{\tilde{C}}^{d-2}(\boldsymbol{\pi}_t \otimes \boldsymbol{Y}_{t+1}) \right\} \\
&= \boldsymbol{\pi}_{t+d-1} \otimes \tilde{\boldsymbol{\omega}} + \mathbb{P}_{\tilde{C}} \tilde{\boldsymbol{Y}}_t(d-1). \qquad (24)
\end{aligned}$$

Equations (23) and (24) provide a convenient scheme for calculation of covariance matrix forecasts.

3.4 Maximum likelihood estimator

In this section, suppose that $\{\boldsymbol{\xi}_t\}$ is normally distributed with mean zero and unit variance matrix. The unknown parameter vector is denoted by $\boldsymbol{\theta}$ and consists of the elements of the transition matrix (2), the regime–specific pairwise correlations $\rho_{\ell m,j}$, $\ell, m = 1, \ldots, M$ ($\ell \neq m$), $j \in \mathscr{E}$, and the parameters of the regime–specific GARCH equations (5); see also Appendix B. Let $\boldsymbol{\theta}_0$ be the true parameter. We assume that the parameter space Θ is a compact subspace of the Euclidean space, such that $\boldsymbol{\theta}_0$ is an interior point in Θ and $\varrho(\mathbb{P}_{C_1}) < 1$ for any $\boldsymbol{\theta} \in \Theta$. The matrix norm is defined by $\|A\| = \sum_{i,j} |A(i,j)|$, where $A(i,j)$ denotes the generic element of matrix A.

Suppose we have a sample of size n, $(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n)$. The estimator of the parameters in model (1)–(7) is obtained by maximizing, conditional on $(\bar{\boldsymbol{\epsilon}}_0, \bar{\boldsymbol{\sigma}}_{10}, \ldots, \bar{\boldsymbol{\sigma}}_{k0})$,

$$L_n(\boldsymbol{\theta}) = L_n(\boldsymbol{\theta}, \boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n) = \sum_{(e_1,\ldots,e_n) \in \mathscr{E}^n} \pi_{e_1,\infty} \left\{ \prod_{t=2}^{n} p_{e_{t-1},e_t} \right\} \left\{ \prod_{t=1}^{n} f_{e_t}(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) \right\},$$

where

$$f_{e_t}(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) = \frac{1}{(2\pi)^{M/2} |\boldsymbol{D}_{e_t,t} \boldsymbol{R}_{e_t} \boldsymbol{D}_{e_t,t}|^{1/2}} \exp\left\{ -\frac{1}{2} \boldsymbol{\epsilon}_t' (\boldsymbol{D}_{e_t,t} \boldsymbol{R}_{e_t} \boldsymbol{D}_{e_t,t})^{-1} \boldsymbol{\epsilon}_t \right\},$$

and the $\boldsymbol{\sigma}_{e_t,t} = (\sigma_{1,e_t,t}, \ldots, \sigma_{M,e_t,t})'$, are defined recursively by (5)–(6), i.e.,

$$\sigma_{\ell,e_t,t} = \omega_{\ell,e_t} + \sum_{m=1}^{M} a_{\ell m,e_t}(|\epsilon_{m,t-1}| - \gamma_{\ell m,e_t} \epsilon_{m,t-1}) + \sum_{m=1}^{M} b_{\ell m,e_t} \sigma_{m,e_t,t-1}, \quad \ell = 1, \ldots, M,$$

for $t = 2, \ldots, n$, and

$$\sigma_{\ell,e_1,1} = \omega_{\ell,e_1} + \sum_{m=1}^{M} a_{\ell m,e_1}(|\bar{\epsilon}_{m0}| - \gamma_{\ell m,e_1} \bar{\epsilon}_{m0}) + \sum_{m=1}^{M} b_{\ell m,e_1} \bar{\sigma}_{m,e_1,0}, \quad \ell = 1, \ldots, M.$$

Write $\mathbf{1}_k = (1, \ldots, 1)' \in \mathbb{R}^k$, $h(\boldsymbol{\epsilon}_1) = (\pi_{1,\infty} f_1(\boldsymbol{\epsilon}_1), \ldots, \pi_{k,\infty} f_k(\boldsymbol{\epsilon}_1))'$, and

$$M_\theta(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) = \begin{pmatrix} p_{11} f_1(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) & p_{21} f_1(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) & \cdots & p_{k1} f_1(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) \\ p_{12} f_2(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) & p_{22} f_2(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) & \cdots & p_{k2} f_2(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) \\ \vdots & \vdots & \ddots & \vdots \\ p_{1k} f_k(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) & p_{2k} f_k(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) & \cdots & p_{kk} f_k(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) \end{pmatrix}.$$

Then $L_n(\boldsymbol{\theta})$ can be rewritten as

$$L_n(\boldsymbol{\theta}) = \mathbf{1}_k' \left\{ \prod_{t=2}^{n} M_\theta(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_{n-t+2}) \right\} h(\boldsymbol{\epsilon}_1).$$

Next, the unobserved likelihood function conditional on $(\boldsymbol{\epsilon}_0, \boldsymbol{\epsilon}_{-1}, \ldots)$ is

$$\begin{aligned} \tilde{L}_n(\boldsymbol{\theta}) = \tilde{L}_n(\theta, \boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n) &= \sum_{(e_1, \ldots, e_n) \in \mathscr{E}^n} \tilde{\pi}_1(e_1) \left\{ \prod_{t=2}^{n} p_{e_{t-1}, e_t} \right\} \left\{ \prod_{t=1}^{n} \tilde{f}_{e_t}(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) \right\} \\ &= \prod_{t=1}^{n} g_\theta(\boldsymbol{\epsilon}_t | \boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots), \end{aligned}$$

where

$$\tilde{f}_{e_t}(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) = \frac{1}{(2\pi)^{M/2} |\tilde{\boldsymbol{D}}_{e_t,t} \boldsymbol{R}_{e_t} \tilde{\boldsymbol{D}}_{e_t,t}|^{1/2}} \exp\left\{ -\frac{1}{2} \boldsymbol{\epsilon}_t' (\tilde{\boldsymbol{D}}_{e_t,t} \boldsymbol{R}_{e_t} \tilde{\boldsymbol{D}}_{e_t,t})^{-1} \boldsymbol{\epsilon}_t \right\},$$

the $\tilde{\boldsymbol{\sigma}}_{e_t,t} = (\tilde{\sigma}_{1,e_t,t}, \ldots, \tilde{\sigma}_{M,e_t,t})'$ are given by

$$\tilde{\boldsymbol{\sigma}}_{e_t,t} = \sum_{i=0}^{\infty} \boldsymbol{B}_{e_t}^i \left[\boldsymbol{\omega}_{e_t} + \boldsymbol{A}_{e_t} |\boldsymbol{\epsilon}_{t-1-i}| - (\boldsymbol{A}_{e_t} \odot \boldsymbol{\Gamma}_{e_t}) \boldsymbol{\epsilon}_{t-1-i} \right],$$

$\tilde{\pi}_t(e_t) = P_\theta(\Delta_t = e_t | \boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots)$, and $g_\theta(\boldsymbol{\epsilon}_t | \boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots)$ denotes the conditional density of $\boldsymbol{\epsilon}_t$ given the σ-field generated by $\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots$.

Because the indices of the states of the Markov chain can be permuted without changing the law of the model, the parameters are not strictly identifiable up to permutation. Moreover, conditions have to be assumed which guarantee that there is no $\boldsymbol{\theta} \in \Theta$, $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, which gives rise, in any of the regimes, to the same sequence of conditional covariance matrices as does $\boldsymbol{\theta}_0$; conditions to assure this are discussed in Jeantheau (1998) and Francq and Zakoïan (2012). We introduce the following assumption of identifiability.

Assumption 1. For any $\boldsymbol{\theta} \in \Theta$, if $g_\theta(\boldsymbol{\epsilon}_t | \boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots) = g_{\theta_0}(\boldsymbol{\epsilon}_t | \boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots)$, P_{θ_0}-a.s., then $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Theorem 3. Denote $(\hat{\boldsymbol{\theta}}_n)$ as the solution to $\sup_{\theta \in \Theta} L_n(\boldsymbol{\theta})$. Under Assumption 1, $\hat{\boldsymbol{\theta}}_n \to_{a.s.} \boldsymbol{\theta}_0$.

Table 1: Properties of weekly global stock market and real estate equity returns

	mean	covariance/ correlation matrix		skewness	kurtosis	JB	ARCH(10)
MSCI	0.064	5.077	0.795	−0.762	7.49	1062.7^{***}	175.6^{***}
EPRA/NAREIT	0.044	4.639	6.714	−1.034	10.81	3091.6^{***}	290.0^{***}

The top right entry of the "covariance/correlation matrix" is the correlation coefficient, and the bottom left entry is the covariance. The return vector at time t is $\boldsymbol{r}_t = (r_{1t}, r_{2t})'$, where r_{1t} and r_{2t} are the MSCI and FTSE EPRA/NAREIT returns, respectively, i.e., the first asset is the MSCI world stock index. JB is the Jarque–Bera test for normality, and ARCH(10) is the LM test for ARCH effects with 10 lags (cf. Engle, 1982). Asterisks *** indicate significance at the 1% level.

4 Application to financial data

To illustrate the theory developed in the previous sections, we consider volatility and correlation dynamics of global stock market and real estate equity returns, using dollar–denominated weekly (Wednesday-to-Wednesday) returns of the MSCI world and the FTSE EPRA/NAREIT global indices from January 1990 to October 2011 ($T = 1137$ observations), with the latter index representing the evolution of real estate equities.[3] The analysis is based on continuously compounded percentage returns, i.e., $r_{it} = 100 \times \log(I_{it}/I_{i,t-1})$, $i = 1, 2$, where I_{1t} and I_{2t} are the MSCI and the FTSE EPRA/NAREIT index levels, respectively.[4] Both the index levels and the returns are shown in the top and middle panels of Figure 1, reflecting the turbulent development of markets particularly since the beginning of the current millennium. Sample moments of the returns are reported in Table 1, along with the Jarque–Bera (JB) test for normality and Engle's (1982) Lagrange Multiplier (LM) test for conditional heteroskedasticity.

From Table 1, we note that the return series exhibit a considerable correlation of 0.795, which reflects the common finding that real estate equities display much more similarity to the general stock market than direct real estate investments (e.g., Morawski et al, 2008; Heaney and Srinanthakumar, 2012). Moreover, the bottom panel of Figure 1 shows conditional correlations implied by an exponentially weighted moving average (EWMA) estimator (cf. Alexander, 2008, Ch. 3.8), which hints at time–varying conditional correlations with a particularly strong degree of comovement both at the beginning and the end of the sample, with the latter being also characterized by an outburst of unprecedented volatility.[5] Results for versions of Tse's

[3] Over a slightly shorter time span, these indices were analyzed in Haas (2010) who shows that Pelletier's (2006) model improves global minimum variance portfolios relative to the standard CCC model.

[4] Data and `Matlab` code for all computations in this section are available from the authors.

[5] Strong evidence for time–varying conditional correlations between these markets has recently been reported by Lee (2014).

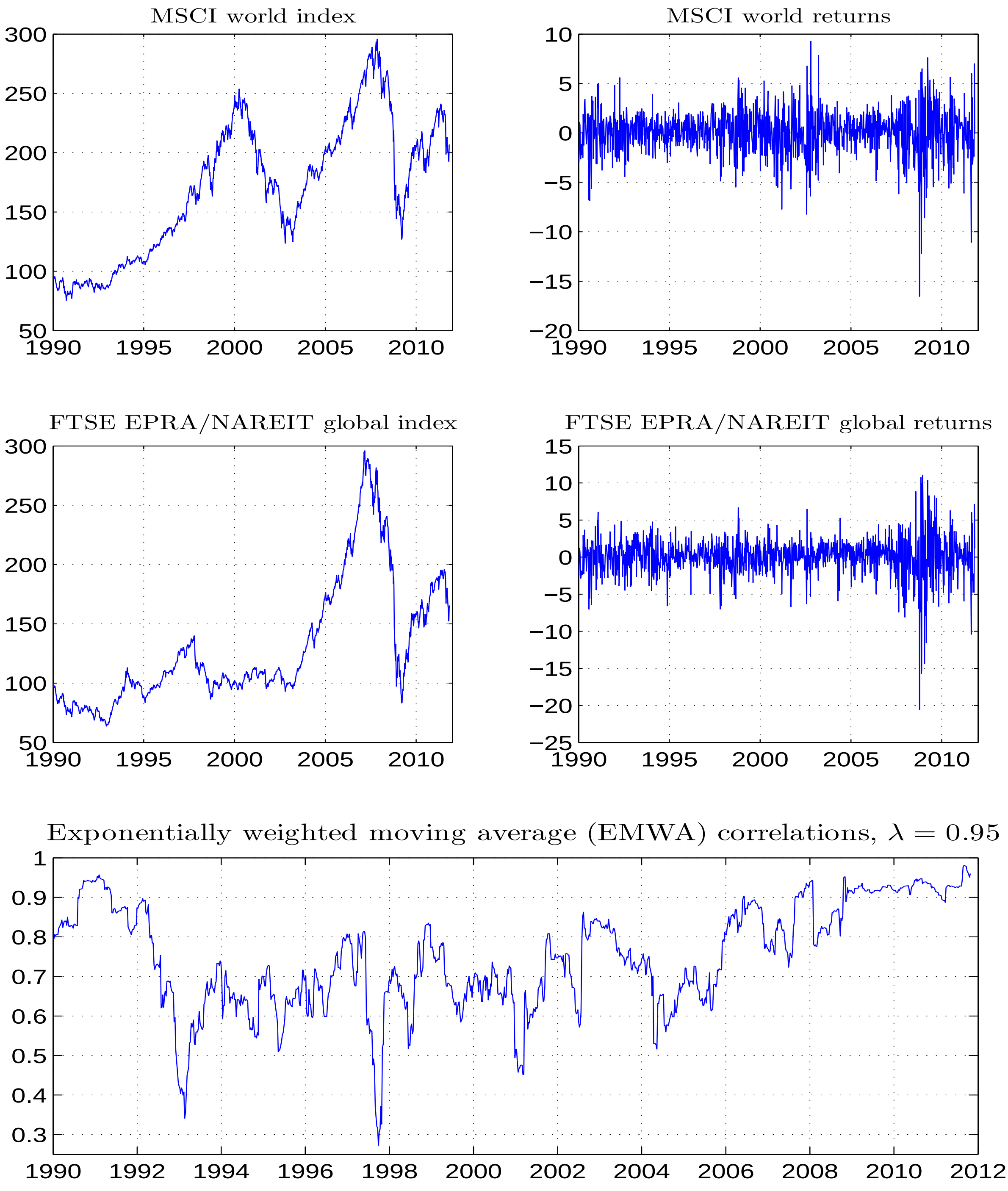


Figure 1: The top panel shows the weekly index levels (left plot) and percentage log returns (right plot) of the MSCI world stock market index from January 1990 to October 2011. The middle panel is similar, but for the FTSE EPRA/NAREIT global index reflecting the evoluation of real estate equities. The bottom panel shows conditional correlations implied by an exponentially weighted moving average (EWMA) covariance matrix estimator $\boldsymbol{H}_t$ with smoothing constant $\lambda = 0.95$, i.e., $\boldsymbol{H}_t = (1-\lambda)\boldsymbol{r}_{t-1}\boldsymbol{r}'_{t-1} + \lambda \boldsymbol{H}_{t-1}$, where the initial matrix $\boldsymbol{H}_1$ is set equal to the sample covariance matrix.

Table 2: Tse's (2000) test for constant conditional correlations

	Gaussian innovations	Student's t innovations
test statistic	7.16^{***}	9.84^{***}

Reported are the results of Tse's (2000) Lagrange Multiplier (LM) test for constant conditional correlations und the assumption of both Gaussian (left column) and Student's t (right column) innovations. Under the null hypothesis, returns are generated by a CCC–AGARCH process as in (B.1), with $k = 1$ and (B.3) imposed in (B.2). Under the null, the LM test statistic given by (B.18) has a limiting $\chi^2(1)$ distribution. Asterisks *** indicate significance at the 1% level.

(2000) Lagrange Multiplier (LM) test for constant conditional correlations in a multivariate GARCH model are reported in Table 2 and also provide support for time–varying conditional correlations.

4.1 Fitting MS CCC–GARCH(1,1) processes

The evidence for time–varying correlations coupled with periods of low and high volatility makes the MS CCC–GARCH model defined in Section 2 a candidate for modeling these series. We fit the model with $k = 1, 2$, and 3 regimes,[6] where we confine ourselves to the *diagonal* model with $\boldsymbol{A}_j$, $\boldsymbol{B}_j$, and $\boldsymbol{\Gamma}_j$ in (5) being diagonal matrices, $j = 1, \ldots, k$. In addition, we restrict the asymmetry parameters to be regime–independent, i.e.,

$$\boldsymbol{\Gamma}_1 = \boldsymbol{\Gamma}_2 = \cdots = \boldsymbol{\Gamma}_k =: \boldsymbol{\Gamma}. \tag{25}$$

There are no clear–cut signs of conditional mean dynamics in the data, and thus we specify the model for return vector $\boldsymbol{r}_t$ as

$$\boldsymbol{r}_t = \boldsymbol{\mu} + \boldsymbol{\epsilon}_t, \tag{26}$$

where $\boldsymbol{\mu}$ is the constant conditional mean and $\boldsymbol{\epsilon}_t$ is generated by a MS(k) CCC–GARCH process as described in Section 2. We compare the fit of models with different k by means of the Bayesian information criterion (BIC) of Schwarz (1978), which, from results of Keribin (2000) and Francq et al. (2001), can be expected to have favorable properties for this purpose. Results are reported in Table 3 for both Gaussian and Student's t innovations $\{\boldsymbol{\xi}_t\}$ in (3). In both cases, models with two components are preferred, as is a conditional t distribution. Thus we focus on two–component models in the following discussion. Both normal and Student's t innovations are considered in order to highlight the role of the conditional distribution.

[6] The model with one regime is just the AGARCH version of the single–component CCC.

Table 3: Likelihood–based goodness–of–fit of MS(k) CCC–GARCH models

	Gaussian innovations			Student's t innovations		
	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
K	11	20	31	12	21	32
$\log L$	-4371.5	-4288.0	-4261.0	-4323.7	-4260.0	-4247.6
BIC	8820.4	8716.6	8740.1	8731.7	8667.7	8720.3

Reported are likelihood–based goodness–of–fit measures for *diagonal* MS(k) CCC–GARCH models fitted to the MSCI world and FTSE EPRA/NAREIT global returns. The number of regimes is denoted by k, and $k = 1$ corresponds to the single–regime CCC of Bollerslev (1990). In all models, the asymmetry parameters are restricted to be constant across regimes, i.e., in (5), $\mathbf{\Gamma}_1 = \cdots = \mathbf{\Gamma}_k$. K is number of parameters of a model, $\log L$ is the value of the maximized log–likelihood, and BIC is the Bayesian information criterion of Schwarz (1978), i.e., BIC $= -2 \times \log L + K \log T$, where T is the sample size. Smaller values of BIC are preferred.

The diagonal MS(k) CCC–GARCH model without further restrictions is rather flexible in that it allows the variances as well as the correlations being regime–dependent. The contribution of both of these features to the overall improvement over the single–regime specification documented in Table 3 is not clear *a priori*. It is thus of interest to test various restricted models against the unrestricted specification. Specifically, we consider Pelletier's (2006) RSDC model where the switching applies to the conditional correlation matrix only, i.e., conditional volatilities are constant across regimes. The second constrained specification represents the opposite of Pelletier's (2006) model, namely the case where volatility can switch but $\boldsymbol{R}_1 = \boldsymbol{R}_2$ in (3). The results reported in Table 4 show that, although both restrictions are rejected against the full model by means of likelihood ratio tests, allowance for regime–specific correlations appears to me more important than switching in the univariate GARCH dynamics, and particularly so for the (generally preferred) models with Student's t innovations.

Several characteristics of the estimated MS(2) CCC–GARCH models with Gaussian and Student's t innovations are reported in Table 5, where the single regime CCC–GARCH models have been included for comparison purposes. In Table 5, the regimes have been ordered such that $\pi_{1,\infty} > \pi_{2,\infty}$. Both two regime–models have in common that Regime 1 is a low–volatility regime with moderate correlation (as related to the unconditional correlation) and Regime 2 is a high–volatility regime with rather high correlation, i.e., the diversification potential deteriorates in turbulent market periods. As reported in the bottom part of Table 5, the unconditional moments implied by the single–regime models are close to those of the two–regime specifications and are in between their regime–specific counterparts documented in the

Table 4: Likelihood ratio tests (LRT) of restricted MS(2) CCC–GARCH specifications against the full (diagonal) model

	Gaussian innovations			Student's t innovations		
	full	Pelletier (RSDC)	$\boldsymbol{R}_1 = \boldsymbol{R}_2$	full	Pelletier (RSDC)	$\boldsymbol{R}_1 = \boldsymbol{R}_2$
K	20	14	19	21	15	20
$\log L$	-4288.0	-4311.7	-4313.0	-4260.0	-4272.6	-4310.9
LRT	–	47.5^{***}	50.1^{***}	–	25.2^{***}	101.9^{***}

The table reports likelihood ratio tests (LRT) for restricted versions of the two–component diagonal MS(2) CCC–GARCH model. The unrestricted specification, denoted as "full", is the model introduced in Section 2 with $k = 2$, and where the matrices $\boldsymbol{A}_j$, $\boldsymbol{B}_j$, $j = 1, 2$, and $\boldsymbol{\Gamma}_1 = \boldsymbol{\Gamma}_2$ in (5) are diagonal. "Pelletier" refers to Pelletier's (2006) regime–switching dynamic correlation (RSDC) model where only the correlation matrix is subject to regime–switching, i.e., the additional restrictions $\boldsymbol{\omega}_1 = \boldsymbol{\omega}_2$, $\boldsymbol{A}_1 = \boldsymbol{A}_2$, and $\boldsymbol{B}_1 = \boldsymbol{B}_2$ are imposed in (5). The third model restricts the correlation to be the same in both regimes, i.e., $\boldsymbol{R}_1 = \boldsymbol{R}_2$ in (3). $\log L$ is the value of the maximized log–likelihood, and K is the number of parameters of a model. The associated likelihood ratio test statistics, denoted as LRT, have 6 and 1 degrees of freedom, respectively. Asterisks *** indicate significance at the 1% level.

top and middle parts of the table for Regimes 1 and 2, respectively. All estimated models are covariance stationary, i.e., $\varrho(\mathbb{P}_{C_2}) < 1$ for all estimated specifications (cf. Theorem 2).

Comparing the regime–switching models with Gaussian and Student's t innovations, we observe that both models are characterized by fairly persistent regimes, but the persistence is more pronounced with Student's t innovations, where both "staying probabilities" p_{11} and p_{22} are rather close to unity. To illustrate the differences in estimated persistence, expected regime durations as implied by estimated parameters, given by $(1 - \widehat{p}_{jj})^{-1}$, $j = 1, 2$, are also reported in Table 5. With Gaussian regimes, expected duration of the low (high)–volatility regime is slightly longer (shorter) than one year, whereas it is almost five (four) years with Student's t regimes.[7] This pattern, which is also discussed in Bulla (2011), Haas (2009, 2010), and Haas and Paolella (2012), is due to the tendency of a model with Gaussian regime densities to signal a regime shift whenever an untypically large (small) observation occurs within an otherwise calm (turbulent) regime.[8] Such untypical observations are easier accommodated when the regime densities are leptokurtic, i.e., display fatter tails and higher peaks than the normal.

[7] But note that expected regime durations may be subject to considerable estimation error; see the legend of Table 5.

[8] The effect of "outliers" on parameter estimates in Gaussian Markov–switching models has also already been reported in the important contribution by Ryden et al. (1998).

Table 5: Characteristics of estimated MS CCC–GARCH(1,1) models

estimated characteristic	Gaussian innovations $k=1$	Gaussian innovations $k=2$	Student's t innovations $k=1$	Student's t innovations $k=2$
$\rho_{12,1}$	0.766 (0.013)	0.636 (0.025)	0.769 (0.014)	0.655 (0.024)
$\mathrm{E}(\epsilon_{1t}^2 \mid \Delta = 1)$	4.265	3.691	3.847	3.161
$\mathrm{E}(\epsilon_{2t}^2 \mid \Delta = 1)$	5.528	3.540	4.713	3.318
$\mathrm{Corr}(\epsilon_{1t}, \epsilon_{2t} \mid \Delta_t = 1)$	0.754	0.617	0.756	0.636
p_{11}	1	0.985 (0.007)	1	0.996 (0.003)
$\pi_{1,\infty}$	1	0.610 (0.151)	1	0.563 (0.109)
$(1-p_{11})^{-1}$	∞	66.54^a	∞	233.8^a
$\rho_{12,2}$	–	0.929 (0.009)	–	0.921 (0.009)
$\mathrm{E}(\epsilon_{1t}^2 \mid \Delta = 2)$	–	6.080	–	5.846
$\mathrm{E}(\epsilon_{2t}^2 \mid \Delta = 2)$	–	7.867	–	7.785
$\mathrm{Corr}(\epsilon_{1t}, \epsilon_{2t} \mid \Delta_t = 2)$	–	0.916	–	0.912
p_{22}	0	0.976 (0.012)	0	0.994 (0.005)
$\pi_{2,\infty}$	0	0.390 (0.151)	0	0.437 (0.109)
$(1-p_{22})^{-1}$	–	42.49^a	–	181.8^a
$\mathrm{E}(\epsilon_{1t}^2)$	4.265	4.622	3.847	4.336
$\mathrm{E}(\epsilon_{2t}^2)$	5.528	5.226	4.713	5.272
$\mathrm{Corr}(\epsilon_{1t}, \epsilon_{2t})$	0.754	0.779	0.756	0.805
$\delta = p_{11} + p_{22} - 1$	–	0.961 (0.017)	–	0.990 (0.006)
ν	–	–	7.368 (1.025)	8.584 (1.370)
γ_{11}	0.680 (0.157)	0.738 (0.168)	0.575 (0.181)	0.627 (0.180)
γ_{22}	0.401 (0.105)	0.635 (0.147)	0.273 (0.116)	0.472 (0.152)
$\varrho(\mathbb{P}_{C_2})$	0.903	0.933	0.921	0.953

Standard errors are given in parentheses. $\rho_{12,j}$ is the constant *conditional* correlation in Regime j; the $\pi_{j,\infty}$ are the stationary regime probabilities; and $(1-p_{jj})^{-1}$ is the expected duration of the jth regime, $j = 1, 2$. $\delta = p_{11}+p_{22}-1$ is a measure for the persistence in the regime process; γ_{11} and γ_{22} are the asymmetry parameters in the volatility equation (5) for the MSCI and the FTSE/NAREIT, respectively, and $\varrho(\mathbb{P}_{C_2})$ is the largest eigenvalue of matrix $\mathbb{P}_{C_2}$ defined in (10) and (11) with $l = 2$, and with $\varrho(\mathbb{P}_{C_2}) < 1$ being the condition for covariance stationarity (cf. Theorem 2).

[a] Asymptotic standard errors for the expected duration of regime j, $(1 - p_{jj})^{-1}$, $j = 1, 2$, could, at least in principle, also be calculated via the delta method. However, with $\widehat{p}_{jj}$ rather close to unity, as in the case under consideration, the normal approximation would be basically useless and thus we abstain from reporting them. For example, with the estimates in the table above, the asymptotic standard error of $(1-\widehat{p}_{11})^{-1}$ in the Student's t switching model would be estimated as $\sqrt{\widehat{\mathrm{Var}}(\widehat{p}_{11})}/(1-\widehat{p}_{11})^2 = 0.0028/(1-0.9957)^2 = 151.4$.

The same logic applies to Pelletier's model where only the correlations are subject to regime–switching, since, for fixed correlation, simultaneous extreme realizations of both variables are more likely with Student's t innovations.

The upper panel of Figure 2 illustrates the models' inferred switching activity by means of the smoothed regime probabilities of the high–volatility/correlation regime under both types of innovation distributions. Both models indicate a switch to the high–correlation regime at the end of the sample beginning with the financial turmoil in the wake of the burst of the housing bubble. Implications for forecasting are depicted in the lower panel of Figure 2: The left plot of the lower panel of Figure 2 shows conditional correlations as implied by a Gaussian regime–switching model for the two situations where we know for certain that at the forecast origin we are either in the first or second regime.[9] As a function of the forecast horizon, the conditional correlation of the Gaussian model rapidly converges to its unconditional value, whereas forecasts are much more persistently affected by the current state of the world in the Student's t model, as shown in the bottom right graph of Figure 2.

4.2 Testing for within–regime correlation dynamics and comparison with other models

One of the most popular approaches to time–varying conditional correlations is the dynamic conditional correlation (DCC) model of Engle (2002). In the DCC, conditional correlations are driven by standardized shocks rather than by discrete regime shifts as in the Markov–switching processes studied herein. Both models can be combined to produce an even more flexible structure which allows the conditional correlations in each regime to be driven by DCC–type dynamics (e.g., Billio and Caporin, 2005; Otranto, 2010). However, the MS CCC–GARCH model has several advantages over its DCC–type generalization, since it is easier to estimate and admits the computation of multi–step–ahead conditional covariance matrices. In view of these advantages, it is desirable to have at one's disposal a simple test of the regime–switching CCC against the alternative of within–regime correlation dynamics. To this end, we extend Tse's (2000) Lagrange Multiplier (LM) test for constant conditional correlations to the multi–regime framework and allowing for fat–tailed (Student's t) innovations.[10] The details of this test, which fits into the general framework described by Hamilton (1996), are developed in

[9] The conditional standard deviations have been initialized by appropriate unconditional expectations, i.e., if $\pi_{\ell t} = 1$, then we set $\sigma_{ij,t+1} = \mathrm{E}(\sigma_{ij,t}|\Delta_{t-1} = \ell)$, $i, j, \ell = 1, 2$.

[10] Silvennoinen and Teräsvirta (2009a) follow a similar approach by deriving LM tests for a CCC against a smooth transition conditional correlation model.

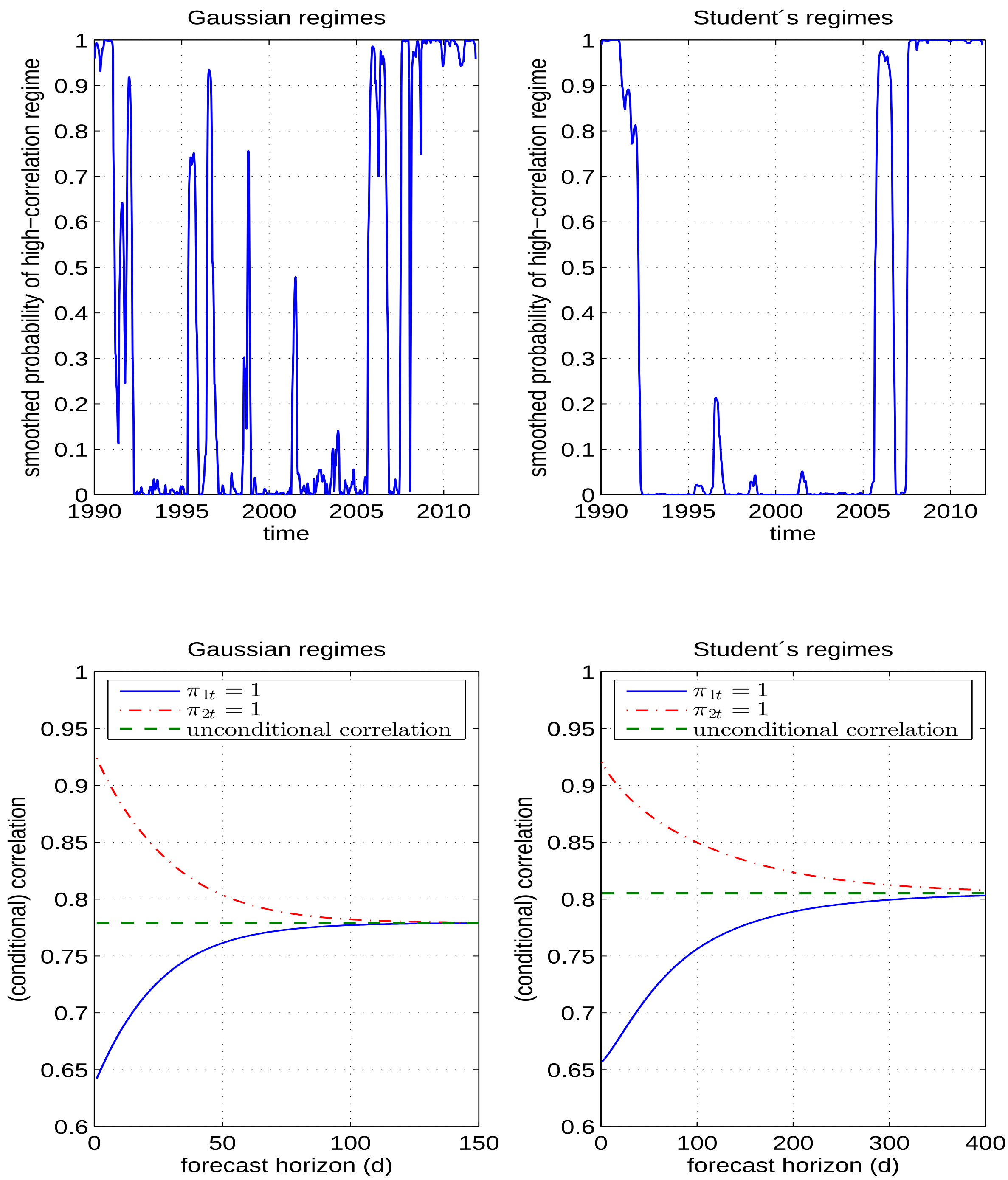


Figure 2: The upper panel shows the smoothed probabilities of Regime 2 (high–volatility/correlation) implied by the MS(2) CCC–GARCH process for Gaussian (left plot) and Student's t innovations (right plot). The lower panel shows conditional correlations under the assumption that we either start in the low– or high–volatility/correlation regime, as represented by the solid and dash–dotted lines, respectively. Conditional standard deviations have been initialized with appropriate unconditional expectations (cf. Footnote 9). As in the upper panel, the left and right graphs are for Gaussian and Student's t innovations, respectively.

Table 6: Lagrange Multiplier (LM) tests for constant within–regime correlations

	Gaussian regime densities	
	specification of alternative hypothesis	
	case (a)	case (b)
Pelletier (RSDC)	4.73*	0.32
MS CCC	0.46	0.33
	Student's t regime densities	
	specification of alternative hypothesis	
	case (a)	case (b)
Pelletier (RSDC)	1.46	0.12
MS CCC	4.48	0.00

Reported are Lagrange Multiplier (LM) tests for constant conditional *within–regime* correlations in two–regime (k) MS CCC–GARCH models, as described in Appendix B. "Pelletier" is Pelletier's (2006) regime–switching dynamic correlation (RSDC) model where only the correlation is subject to regime–switching. Note that the tests reported here are based on symmetric volatility processes, i.e., $\Gamma_1 = \Gamma_1 = \mathbf{0}$ in (5) Cases (a) and (b) are distinguished by means of the alternative hypothesis as described in Appendix B:

- Case (a) refers to the situation where, under the alternative, the correlation dynamics may be different in both regimes, i.e., in (B.2), both $\delta_{12,j}$, $j = 1, 2,$, may be nonzero and different.
- In case (b), it is assumed under the alternative that correlation dynamics are regime–independent, i.e., $\delta_{12,1} = \delta_{12,2}$ in (B.2).

Under the null hypothesis of constant conditional correlations in both regimes, the test statistic for case (a) is asymptotically distributed as $\chi^2(2)$, whereas in case (b) the limiting distribution is $\chi^2(1)$. Asterisk * denotes significance at the 10% level (the p–value is 0.094).

Appendix B. Results are reported in Table 6 for two conditional volatility specifications under the null hypothesis, that is, both the "full" model from Table 4 as well as Pelletier's model, and both specifications are considered with Gaussian and Student's t innovations. Under the alternative, within–regime correlation dynamics are either regime–dependent (case (a) in Table 6) or regime–independent (case (b)). Overall, the results in Table 6 show no clear–cut sign of within–regime correlation dynamics, i.e., the switching between low– and high–correlation periods appears to capture most of the time–variation in conditional correlations.

In view of these results, we compare conditional correlations implied by the MS CCC and the DCC model of Engle (2002).[11] These correlations are shown in Figure 3 for both Gaussian (top panel) and Student's t innovations (bottom panel). Comparing the upper

[11] For purpose of comparison with the other models discussed herein, the DCC is likewise coupled with asymmetric absolute value GARCH for the volatilities.

Table 7: Parameter estimates for correlation dynamics in DCC models

innovations	$\widehat{a}$	$\widehat{b}$	$\widehat{a}+\widehat{b}$
Gaussian	0.044 (0.014)	0.942 (0.022)	0.986 (0.009)
Student's t	0.047 (0.015)	0.941 (0.021)	0.988 (0.008)

Shown are estimates of the parameters driving the correlation dynamics in Gaussian and Student's t DCC models à la Engle (2002), with standard errors given in parentheses. The evolution of the conditional correlation matrix $\boldsymbol{R}_t$ is described by

$$\begin{aligned}\boldsymbol{Q}_t &= (1-a-b)\boldsymbol{S} + a\boldsymbol{z}_{t-1}\boldsymbol{z}'_{t-1} + b\boldsymbol{Q}_{t-1}\\ \boldsymbol{R}_t &= (\boldsymbol{I}\odot\boldsymbol{Q}_t)^{-1/2}\boldsymbol{Q}_t(\boldsymbol{I}\odot\boldsymbol{Q}_t)^{-1/2},\end{aligned}$$

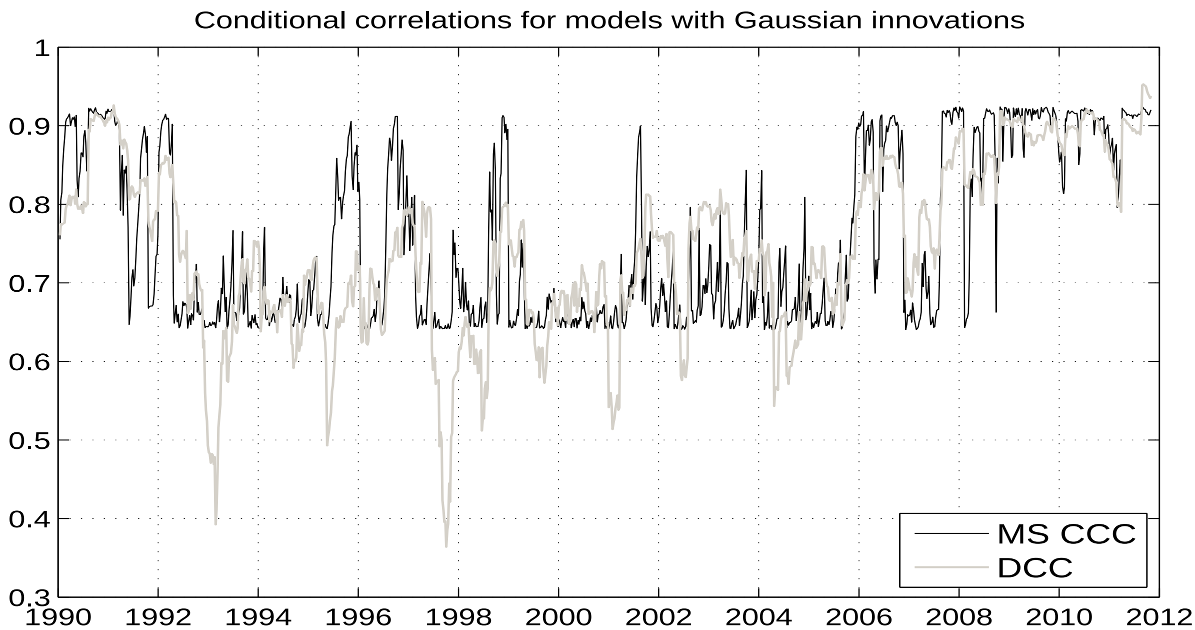
where the $\boldsymbol{z}_t$ are the standardized ("degarched") residuals, and $\boldsymbol{S}$ is estimated via their sample correlation matrix.

with the lower panel of Figure 3, MS CCC-implied correlations are smoother with Student's t then with Gaussian innovations. The DCC–implied correlations depend much less on the innovation distribution and, as already observed by Pelletier (2006), are less smooth than their regime–switching CCC counterparts.[12] Roughly, however, both types of models contain similar information about low– and high–correlation periods in the data. In particular, they agree with regard the jump in correlation at the onset of the recent financial crisis.

Multi–step ahead conditional correlations for both types of models are illustrated in Figure 4 which resembles the lower part of Figure 2 but additionally includes conditional correlations implied by the Student's t DCC model, as calculated by simulation.[13] Initial values for the conditional correlation matrix and the conditional standard deviations in the DCC were selected such that they match those of the Student's t MS CCC in the respective regimes. The long–run correlation of the DCC model is a bit lower than those implied by estimated MS CCC processes. However, with regard to the persistence of multi–step correlations, the DCC is more like the Student's t rather than the Gaussian MS CCC–GARCH. This is in line with DCC parameter estimates, as reported in Table 7. Interpreting $\widehat{a}+\widehat{b}$ as an estimate of the persistence of conditional correlations, i.e., the equivalent of δ in Table 5, then the persistence in conditional correlations implied by estimated DCC models is close to the value of $\widehat{\delta}=0.99$ of the Student's t MS CCC model.

[12] As argued by Pelletier (2006), "[o]ne interesting implication of smoother patterns for the correlations is for the computation of VaR and portfolio allocation. If the time–varying correlations are smoother, then the gain from portfolio diversification will also be smoother which might imply a smoother pattern for the VaR and portfolio weights."

[13] Those for the Gaussian DCC are quite similar and not shown.



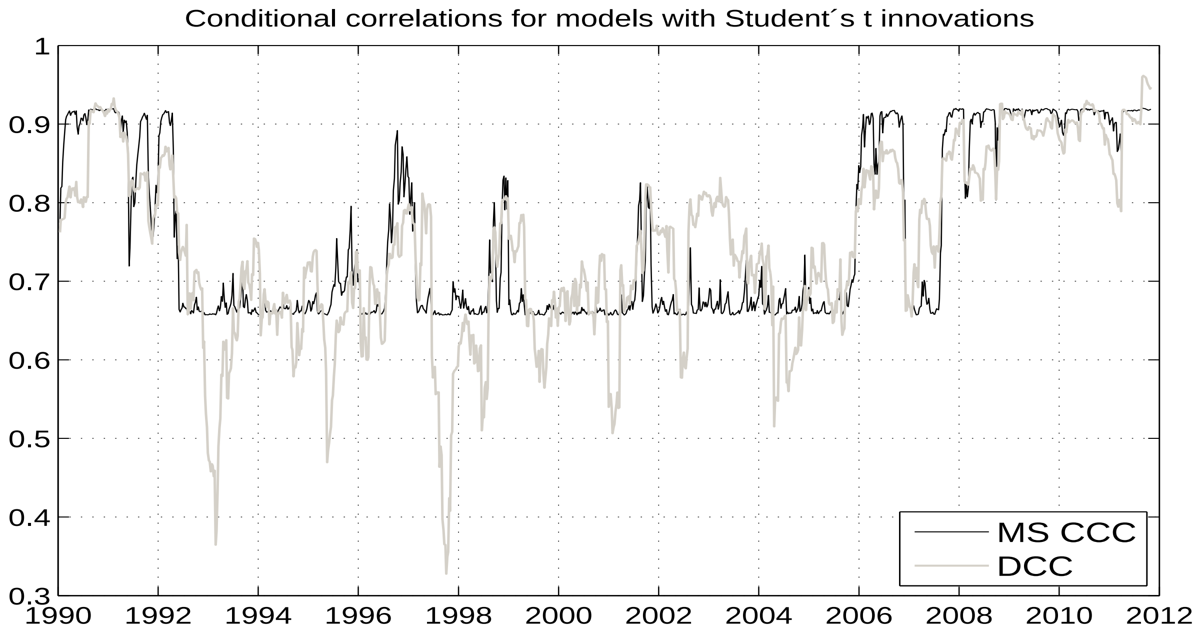


Figure 3: The upper panel show conditional correlations as implied by the MS CCC–GARCH model and the DCC process with Gaussian innovations. The lower panel repeats this, but for models with Student's t innovations.

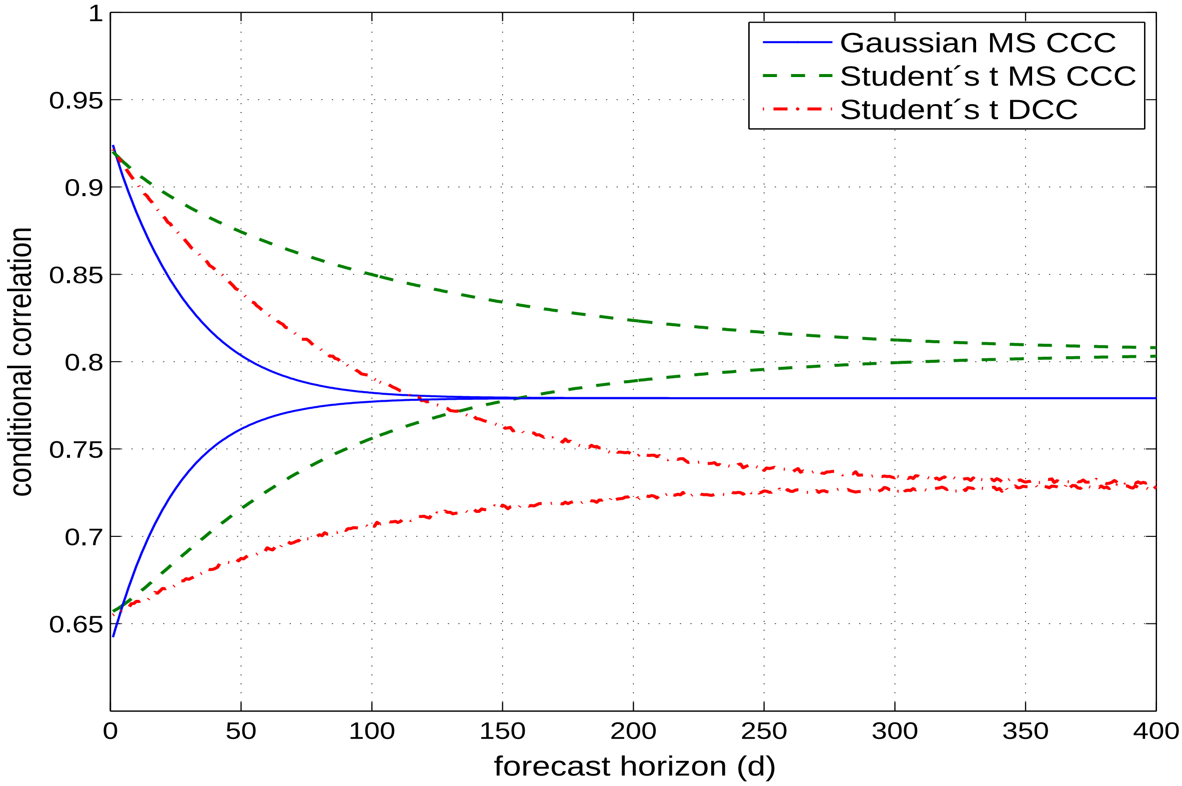


Figure 4: Shown are conditional correlations similar to the lower part of and as explained in the legend of Figure 3. The curves for the MS CCC–GARCH models reproduce those in the bottom part of Figure 3. Initial values for the conditional correlation matrix and the conditional standard deviations in the DCC model were determined such that they match those of the Student's t MS CCC in the respective (low– and high correlation/variance) regimes.

4.3 Application to portfolio selection

We finally compare the models' performance in an out–of–sample portfolio application. To do so, we first reestimate all models using roughly the first ten years of data (i.e., the first 500 observations) and then update the estimates every four weeks, using an expanding window of observations. Estimated models are used to construct *ex–ante* global minimum variance portfolios (GMVP) over holding periods up to 24 weeks (ca. 6 months).[14] Using non–overlapping holding periods, we thus have, e.g., 637 and 318 out–of–sample realized GMVP returns for the one– and two–week holding periods, respectively. Closed–form conditional covariances as developed in Section 3.3 are used for all CCC–type models, whereas the DCC–implied

[14] The GMVP has the advantage that it allows to focus on the covariance matrix without "having to specify the vector of conditional expected returns, which is more a task for the portfolio manager than a statistical problem" (Ledoit et al. 2003).

conditional covariance matrices are estimated from 10,000 simulated sample paths.

Results are reported in Table 8. For the most basic model, i.e., the single–regime Gaussian CCC, we report the standard deviation of the realized returns, whereas for all other models their respective standard deviation divided by that of the Gaussian CCC is shown. The results in Table 8 show that using a Student's t rather than a Gaussian distribution improves the results somewhat for all models and forecast horizons. However, the improvements tend to be minor except for the MS CCC–GARCH for which they are quite substantial, and even more so for longer forecast horizons. At first, it may appear surprising that the MS CCC with Studnet's t innovations displays the best results for all forecast horizons, whereas the performance of its Gaussian cousin is rather disappointing. However, this becomes plausible in view of the discussion in Section 4.1. Namely, in the Gaussian model, with volatilities being allowed to switch, the high–volatility regime will tend to latch onto a few "outliers", which hampers the ability of the model to identify the smooth and long–lived low– and high–correlation regimes. Pelletier's RSDC model, with switching correlations only, does not suffer from this problem, and thus its performance is more even across distributional assumptions. Still, however, the results for the MS CCC with t errors in Table 8 suggest that there may be additional benefits from allowing both correlations and volatilities being regime–dependent, provided the conditional distribution is flexible enough to cope with isolated untypical observations within a given regime. We finally note that also RSDC consistently outperforms DCC at longer forecast horizons, which may indicate that the regime–switching models are better suited to capture relatively long–lived persistent correlation regimes.

5 Conclusion

We want to conclude with two remarks. The first relates to the properties of the MLE considered in Section 3.4. Its consistency was established for Gaussian innovations. In view of the empirical relevance of nonnormal regimes demonstarted in Section 4, it would be interesting to extend the results of Section 3.4 to such less restrictive settings. However, we believe that the result given in Section 3.4 is of some interest since to our knowledge no such results even for the univariate model of Haas et al. (2004) have been available so far.

The second remark refers to the frequently contemplated "curse of dimensionality" problem. An advantage of the (diagonal) CCC, DCC, and Pelletier's (2006) RSDC models is that, via two–step estimation, application to high–dimensional time series is feasible. This property is not shared by the model studied herein with both regime–specific correlations and variance

Table 8: Realized standard deviations of out–of–sample global minimum variance portfolio (GMVP) returns

	Models with Gaussian innovations								
horizon (D)	1	2	3	4	8	12	16	20	24
# returns	637	318	212	159	79	53	39	31	26
CCC	2.425	3.324	4.472	4.926	7.192	9.823	10.791	15.506	17.439
DCC	0.954	0.986	0.960	0.979	0.970	0.961	0.992	0.986	0.978
RSDC	0.971	0.973	0.965	0.962	0.952	0.941	0.966	0.942	0.926
MS CCC	0.982	0.989	0.986	0.995	1.010	0.984	1.046	1.040	1.001
	Models with Student's t innovations								
horizon (D)	1	2	3	4	8	12	16	20	24
# returns	637	318	212	159	79	53	39	31	26
CCC	0.996	0.995	0.993	0.991	0.991	0.979	0.992	0.990	0.981
DCC	0.946	0.981	0.945	0.968	0.956	0.928	0.978	0.963	0.947
RSDC	0.961	0.968	0.951	0.953	0.935	0.901	0.952	0.924	0.895
MS CCC	0.936	0.965	0.924	0.938	0.902	0.862	0.894	0.843	0.821

Reported are the results of constructing *ex–ante* global minimum variance portfolios (GMVP) implied by different GARCH models and for different forecast horizons, D (weeks). Calculations refer to multi–period cumulative returns, i.e., if $\boldsymbol{r}_{t+d}$ is the single–period return vector at time (week) $t+d$, then the D–period ahead cumulative return vector at forecast origin t is $\sum_{d=1}^{D} \boldsymbol{r}_{t+d}$, and the multi–period ahead covariance matrices are calculated accordingly (assuming returns are not autocorrelated in the current application). The row labeled "# returns" reports the number of (non–overlapping) holding periods used to produce the results for each respective forecast horizon. CCC and DCC are Bollerslev's (1990) constant and Engle's (2002) dynamic conditional correlation models, respectively. RSDC is Pelletier's (2006) model, and MS DCC is the Markov–switching GARCH process defined in Section 2. In all models, volatilities are driven by absolute value asymmetric GARCH processes, cf. Equation (5).

For the CCC with normal innovations, the table reports the standard deviation of the *ex–post* (realized) portfolio returns of the *ex–ante* GMVP for each forecast horizon, D. For all other models, their respective standard deviation divided by that of the Gaussian CCC is shown.

dynamics. We do not deem this to be a disadvantage, since there are many applications where the advantage of a more flexible dynamic structure may very well outweigh the benefits of parsimony *as long as the dimensionality of the problem is low to moderate.* Studies of the dynamics of broadly defined asset classes, as illustrated in Section 4, are a typical example, where a richer specification can lead to a better understanding and potentially improved forecasts of the joint process under study. As another recent example from the literature somewhat related to the application in Section 4, Case et al. (2014), using monthly data from 1972 to 2009, find that even a four–regime MS model is required to appropriately describe the evolution of the joint conditional distribution of REIT, stock, and bond returns, since in particular the bond market regimes fail to be synchronized with those of the other two markets. For higher–dimensional systems, Pelletier's (2006) model, which is nested in the general specification of this paper, appears to provide a reasonable balance between flexibility on the one hand and parsimony and tractability on the other. This is suggested in particular sice the results in Section 4 (cf. Table 4) revealed that allowing for regime–switching correlations is of greater value than doing the same for the dynamics of individual volatilities.

Appendix

A Proofs of the theorems

Proof of Theorem 1. The 'if' part follows from Brandt (1986) or Bougerol and Picard (1992).

Conversely, assume that there exists a strictly stationary solution $(\boldsymbol{\epsilon}_t)$ of the MS(k)-CCC-GARCH process defined by (1)–(7). Iterating (8), we have, for any $m > 0$, that

$$\boldsymbol{X}_0 = \boldsymbol{\omega} + \sum_{n=1}^{m} \boldsymbol{C}_{\Delta_{-1},-1}\boldsymbol{C}_{\Delta_{-2},-2}\ldots\boldsymbol{C}_{\Delta_{-n},-n}\boldsymbol{\omega} + \boldsymbol{C}_{\Delta_{-1},-1}\boldsymbol{C}_{\Delta_{-2},-2}\ldots\boldsymbol{C}_{\Delta_{-m-1},-m-1}\boldsymbol{X}_{-m-1}.$$

From all entries of $\boldsymbol{X}_t$, $\boldsymbol{C}_{\Delta_t,t}$, and $\boldsymbol{\omega}$ being nonnegative, we know, for any $m > 0$,

$$\sum_{n=1}^{m} \boldsymbol{C}_{\Delta_{-1},-1}\boldsymbol{C}_{\Delta_{-2},-2}\ldots\boldsymbol{C}_{\Delta_{-n},-n}\boldsymbol{\omega} \le \boldsymbol{X}_0, \quad \text{a.s.}$$

Therefore, $\sum_{n=1}^{m} \boldsymbol{C}_{\Delta_{-1},-1}\boldsymbol{C}_{\Delta_{-2},-2}\ldots\boldsymbol{C}_{\Delta_{-n},-n}\boldsymbol{\omega}$ converges a.s. Thus, we have that

$$\lim_{n\to\infty} \boldsymbol{C}_{\Delta_{-1},-1}\boldsymbol{C}_{\Delta_{-2},-2}\ldots\boldsymbol{C}_{\Delta_{-n},-n}\boldsymbol{\omega} = 0, \quad \text{a.s.}$$

By $\boldsymbol{\omega} > 0$, we have

$$\lim_{n\to\infty} \boldsymbol{C}_{\Delta_{-1},-1}\boldsymbol{C}_{\Delta_{-2},-2}\ldots\boldsymbol{C}_{\Delta_{-n},-n}e_i^* = 0, \quad \text{a.s.}$$

where (e_i^*) is the canonical basis of $\mathbb{R}^{kM}$. Therefore,

$$\lim_{n\to\infty} \left\|\boldsymbol{C}_{\Delta_{-1},-1}\boldsymbol{C}_{\Delta_{-2},-2}\ldots\boldsymbol{C}_{\Delta_{-n},-n}\right\| = 0, \quad \text{a.s.}$$

Hence, by Lemma 3.4 in Bougerol and Picard (1992), we know that the top Lyapunov exponent associated with the matrices $(\boldsymbol{C}_{\Delta_t,t})$ is strictly negative. This completes the proof of the theorem.

Proof of Theorem 2. Write

$$\boldsymbol{X}_{t,m} = \boldsymbol{C}_{\Delta_{t-1},t-1}\boldsymbol{C}_{\Delta_{t-2},t-2}\ldots\boldsymbol{C}_{\Delta_{t-m},t-m}\boldsymbol{\omega}, \quad m \geq 1,$$

and $\boldsymbol{X}_{t,0} = \boldsymbol{\omega}$. For any vector $\boldsymbol{X}$ such that $\boldsymbol{A}\boldsymbol{X}$ is well defined, we have $(\boldsymbol{A}\boldsymbol{X})^{\otimes l} = \boldsymbol{A}^{\otimes l}\boldsymbol{X}^{\otimes l}$. It follows that

$$\boldsymbol{X}_{t,m}^{\otimes l} = \boldsymbol{C}^{\otimes l}_{\Delta_{t-1},t-1}\boldsymbol{C}^{\otimes l}_{\Delta_{t-2},t-2}\cdots\boldsymbol{C}^{\otimes l}_{\Delta_{t-m},t-m}\boldsymbol{\omega}^{\otimes l}.$$

By Lemma 1 in Francq and Zakoïan (2005), we have

$$\begin{aligned}
E(\boldsymbol{X}_{t,m}^{\otimes l}) &= E\{E(\boldsymbol{C}^{\otimes l}_{\Delta_{t-1},t-1}\boldsymbol{C}^{\otimes l}_{\Delta_{t-2},t-2}\cdots\boldsymbol{C}^{\otimes l}_{\Delta_{t-m},t-m}\boldsymbol{\omega}^{\otimes l}|\Delta_{t-1},\ldots,\Delta_{t-m})\} \\
&= E(C_l(\Delta_{t-1})C_l(\Delta_{t-2})\ldots C_l(\Delta_{t-m}))\boldsymbol{\omega}^{\otimes l} \\
&= \mathbb{I}(\mathbb{P}_{C_l})^m \pi_{\boldsymbol{\omega}^{\otimes l}},
\end{aligned}$$

where $\mathbb{I} = (I_{(kM)^l},\ldots,I_{(kM)^l})$ is a $(kM)^l \times k(kM)^l$ matrix and $\pi_{\boldsymbol{\omega}^{\otimes l}} = (\boldsymbol{\pi}_\infty \otimes \boldsymbol{1}_{k^{l-1}M^l}) \odot \boldsymbol{\omega}^{\otimes l}$. Thus, by $\|A\|\|B\| = \|A\otimes B\| = \|B\otimes A\|$, we have

$$(E\|\boldsymbol{X}_{t,m}\|^l)^{1/l} = (\|E(\boldsymbol{X}_{t,m}^{\otimes l})\|)^{1/l} = \|\mathbb{I}(\mathbb{P}_{C_l})^m\pi_{\boldsymbol{\omega}^{\otimes l}}\|^{1/l} \leq \|\mathbb{I}\|^{1/l}\|(\mathbb{P}_{C_l})^m\|^{1/l}\|\pi_{\boldsymbol{\omega}^{\otimes l}}\|^{1/l} \to 0$$

at an exponential rate as $m \to \infty$, because $\varrho(\mathbb{P}_{C_l}) < 1$. This shows that

$$\lim_{n\to\infty}\sum_{m=0}^{n}\boldsymbol{X}_{t,m} = \boldsymbol{X}_t^* = \boldsymbol{\omega} + \sum_{n=1}^{\infty}\boldsymbol{C}_{\Delta_{t-1},t-1}\boldsymbol{C}_{\Delta_{t-2},t-2}\ldots\boldsymbol{C}_{\Delta_{t-n},t-n}\boldsymbol{\omega},$$

both in L^l and almost surely. It is obvious that $\boldsymbol{X}_t^*$ satisfies (8) and is strictly stationary and ergodic. This completes the proof of the theorem.

The proof of Theorem 3 adopts some elements of the method of Francq et al. (2001). The argument is developed along a set of intermediate results which are presented in the following lemmas.

Lemma 2. For all $i \in \mathscr{E}$ and $\boldsymbol{\theta} \in \Theta$, we have

$$\tilde{\pi}_t(i) = P_\theta(\Delta_t = i | \boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots) > 0 \quad P_{\theta_0} - \text{a.s.}$$

Proof. The irreductibility and aperiodicity assumptions imply that (Δ_t) is primitive (see Seneta, 1981, p. 21), i.e. there exists a strictly positive integer m such that

$$P_\theta(\Delta_t = i | \Delta_{t-m} = j) > 0,$$

for any $(i, j) \in \mathscr{E}^2$. Moreover, note that the conditional densities $g_\theta(\boldsymbol{\epsilon}_{t-1}, \ldots, \boldsymbol{\epsilon}_{t-m} | \Delta_{t-m} = j)$ and $g_\theta(\boldsymbol{\epsilon}_{t-1}, \ldots, \boldsymbol{\epsilon}_{t-m} | \Delta_t = i, \Delta_{t-m} = j)$ are almost surely strictly positive. Hence, for all $i \in \mathscr{E}$, we have

$$\begin{aligned}
& P_\theta(\Delta_t = i | \boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots) \\
= & \sum_{j=1}^{k} P_\theta(\Delta_t = i | \Delta_{t-m} = j, \boldsymbol{\epsilon}_{t-1}, \ldots, \boldsymbol{\epsilon}_{t-m}) P_\theta(\Delta_{t-m} = j | \boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots) \\
\geq & \min_{j \in \mathscr{E}} P_\theta(\Delta_t = i | \Delta_{t-m} = j, \boldsymbol{\epsilon}_{t-1}, \ldots, \boldsymbol{\epsilon}_{t-m}) \\
= & \min_{j \in \mathscr{E}} \frac{g_\theta(\boldsymbol{\epsilon}_{t-1}, \ldots, \boldsymbol{\epsilon}_{t-m} | \Delta_t = i, \Delta_{t-m} = j) P_\theta(\Delta_t = i | \Delta_{t-m} = j)}{g_\theta(\boldsymbol{\epsilon}_{t-1}, \ldots, \boldsymbol{\epsilon}_{t-m} | \Delta_{t-m} = j)} > 0.
\end{aligned}$$

This completes the proof of the lemma.

Lemma 3. For all $\boldsymbol{\theta} \in \Theta$,

$$\lim_{n \to \infty} \frac{1}{n} \log L_n(\boldsymbol{\theta}) = \lim_{n \to \infty} \frac{1}{n} \log \tilde{L}_n(\boldsymbol{\theta}) = E_{\theta_0} g_\theta(\boldsymbol{\epsilon}_t | \boldsymbol{\epsilon}_{t-1}, \ldots).$$

Proof. Let $\tilde{h}(\boldsymbol{\epsilon}_1) = (\tilde{\pi}_1(1)\tilde{f}_1(\boldsymbol{\epsilon}_1), \ldots, \tilde{\pi}_1(k)\tilde{f}_k(\boldsymbol{\epsilon}_1))'$ and

$$\tilde{M}_\theta(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) = \begin{pmatrix} p_{11}\tilde{f}_1(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) & p_{21}\tilde{f}_1(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) & \cdots & p_{k1}\tilde{f}_1(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) \\ p_{12}\tilde{f}_2(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) & p_{22}\tilde{f}_2(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) & \cdots & p_{k2}\tilde{f}_2(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) \\ \vdots & \vdots & \ddots & \vdots \\ p_{1k}\tilde{f}_k(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) & p_{2k}\tilde{f}_k(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) & \cdots & p_{kk}\tilde{f}_k(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) \end{pmatrix}.$$

Then $\tilde{L}_n(\boldsymbol{\theta})$ can be rewritten as

$$\tilde{L}_n(\boldsymbol{\theta}) = \mathbf{1}_k' \left\{ \prod_{t=2}^{n} \tilde{M}_\theta(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_{n-t+2}) \right\} \tilde{h}(\boldsymbol{\epsilon}_1).$$

Moreover, we have

$$\begin{aligned}
&\min_j \tilde{\pi}_1(j)\tilde{f}_j(\boldsymbol{\epsilon}_1)\left\|\prod_{t=2}^{n}\tilde{M}_\theta(\boldsymbol{\epsilon}_1,\ldots,\boldsymbol{\epsilon}_{n-t+2})\right\| \\
\leq\ & \tilde{L}_n(\boldsymbol{\theta}) \leq \max_j \tilde{\pi}_1(j)\tilde{f}_j(\boldsymbol{\epsilon}_1)\left\|\prod_{t=2}^{n}\tilde{M}_\theta(\boldsymbol{\epsilon}_1,\ldots,\boldsymbol{\epsilon}_{n-t+2})\right\|.
\end{aligned}$$

By $\varrho(\mathbb{P}_{C_1}) < 1$, it is easy to prove that $E_{\theta_0}|\log g_\theta(\boldsymbol{\epsilon}_t|\boldsymbol{\epsilon}_{t-1},\boldsymbol{\epsilon}_{t-2},\ldots)| < \infty$. Therefore, by Lemma 2, the ergodic theorem implies that

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n}\log\left\|\prod_{t=2}^{n}\tilde{M}_\theta(\boldsymbol{\epsilon}_1,\ldots,\boldsymbol{\epsilon}_{n-t+2})\right\| \\
=\ & \lim_{n\to\infty}\frac{1}{n}\log\tilde{L}_n(\boldsymbol{\theta}) \\
=\ & E_{\theta_0}\log g_\theta(\boldsymbol{\epsilon}_t|\boldsymbol{\epsilon}_{t-1},\boldsymbol{\epsilon}_{t-2},\ldots).
\end{aligned}$$

For convenience, let the initial values be $\bar{\boldsymbol{\epsilon}}_0 = 0, \bar{\boldsymbol{\sigma}}_{j0} = \boldsymbol{\omega}_j, j \in \mathscr{E}$. Then, for $j \in \mathscr{E}$, we have

$$\|\tilde{\boldsymbol{\sigma}}_{j,t} - \boldsymbol{\sigma}_{j,t}\| \leq O(c^t)Q_{1t} := \sum_{i=t}^{\infty} c^i\left[c_1 + c_2|||\boldsymbol{\epsilon}_{t-i}|||\right] = c^t\sum_{i=0}^{\infty} c^i\left[c_1 + c_2|||\boldsymbol{\epsilon}_{-i}|||\right],$$

where $0 < c < 1$ and c_1 and c_2 are constants independent of the parameter $\boldsymbol{\theta}$.

Hence, for $j \in \mathscr{E}$,

$$\begin{aligned}
&|\log|\tilde{\boldsymbol{D}}_{jt}\boldsymbol{R}_j\tilde{\boldsymbol{D}}_{jt}| - \log|\boldsymbol{D}_{jt}\boldsymbol{R}_j\boldsymbol{D}_{jt}|| = \left|\sum_{\ell=1}^{M}\log\left(\frac{\tilde{\sigma}_{\ell jt}^2}{\sigma_{\ell jt}^2}\right)\right| \\
\leq\ & 2\sum_{\ell=1}^{M}\left|\frac{\tilde{\sigma}_{\ell jt} - \sigma_{\ell jt}}{\sigma_{\ell jt}}\right| = O(1)\sum_{\ell=1}^{M}|\tilde{\sigma}_{\ell jt} - \sigma_{\ell jt}| \leq O(c^t)Q_{1t}
\end{aligned}$$

and

$$\begin{aligned}
&|\boldsymbol{\epsilon}_t'(\tilde{\boldsymbol{D}}_{jt}\boldsymbol{R}_j\tilde{\boldsymbol{D}}_{jt})^{-1}\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_t'(\boldsymbol{D}_{jt}\boldsymbol{R}_j\boldsymbol{D}_{jt})^{-1}\boldsymbol{\epsilon}_t| \\
=\ & |\boldsymbol{\epsilon}_t'\tilde{\boldsymbol{D}}_{jt}^{-1}\boldsymbol{R}_j^{-1}(\tilde{\boldsymbol{D}}_{jt}^{-1}\boldsymbol{\epsilon}_t - \boldsymbol{D}_{jt}^{-1}\boldsymbol{\epsilon}_t) - \boldsymbol{\epsilon}_t'\boldsymbol{D}_{j,t}^{-1}\boldsymbol{R}_j^{-1}(\boldsymbol{D}_{jt}^{-1}\boldsymbol{\epsilon}_t - \tilde{\boldsymbol{D}}_{jt}^{-1}\boldsymbol{\epsilon}_t)| \\
\leq\ & O(1)\|\boldsymbol{\epsilon}_t\|\sum_{\ell=1}^{M}\left|\frac{\epsilon_{\ell t}}{\tilde{\sigma}_{\ell jt}} - \frac{\epsilon_{\ell t}}{\sigma_{\ell jt}}\right| + O(1)\|\boldsymbol{\epsilon}_t\|\sum_{\ell=1}^{M}\left|\frac{\epsilon_{\ell t}}{\sigma_{\ell jt}} - \frac{\epsilon_{\ell t}}{\tilde{\sigma}_{\ell jt}}\right| \\
=\ & O(1)\|\boldsymbol{\epsilon}_t\|\sum_{\ell=1}^{M}|\epsilon_{\ell t}|\left|\frac{\tilde{\sigma}_{\ell jt} - \sigma_{\ell jt}}{\tilde{\sigma}_{\ell jt}\sigma_{\ell jt}}\right| \\
\leq\ & O(1)\|\boldsymbol{\epsilon}_t\|^2\sum_{\ell=1}^{M}|\tilde{\sigma}_{\ell jt} - \sigma_{\ell jt}| \leq O(c^t)Q_{2t},
\end{aligned}$$

where $Q_{2t} = \|\boldsymbol{\epsilon}_t\|^2 Q_{1t}$. This shows that, for $j \in \mathscr{E}$,

$$|\log f_j(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) - \log \tilde{f}_j(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t)| \leq O(c^t) Q_t,$$

where $Q_t = Q_{1t} + Q_{2t}$ is stationary and ergodic. By the above inequality, we may get that, for $j \in \mathscr{E}$,

$$\log f_j(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) \leq \log \tilde{f}_j(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) + O(c^t) Q_t,$$

that is,

$$f_j(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t) \leq e^{O(c^t) Q_t} \tilde{f}_j(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t).$$

Then

$$\left\| \prod_{t=2}^{n} M_\theta(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_{n-t+2}) \right\| \leq e^{\sum_{t=2}^{n} O(c^t) Q_t} \left\| \prod_{t=2}^{n} \tilde{M}_\theta(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_{n-t+2}) \right\|.$$

It follows that

$$\begin{aligned}
& \limsup_{n\to\infty} \frac{1}{n} \log \left\| \prod_{t=2}^{n} M_\theta(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_{n-t+2}) \right\| \\
\leq & \lim_{n\to\infty} \frac{1}{n} \log \left\| \prod_{t=2}^{n} \tilde{M}_\theta(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_{n-t+2}) \right\| + \lim_{n\to\infty} \frac{1}{n} \sum_{t=2}^{n} O(c^t) Q_t \\
= & \lim_{n\to\infty} \frac{1}{n} \log \tilde{L}_n(\boldsymbol{\theta}) = E_{\theta_0} \log g_\theta(\boldsymbol{\epsilon}_t | \boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots).
\end{aligned}$$

Similarly, we also have that

$$\liminf_{n\to\infty} \frac{1}{n} \log \left\| \prod_{t=2}^{n} M_\theta(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_{n-t+2}) \right\| \geq E_{\theta_0} \log g_\theta(\boldsymbol{\epsilon}_t | \boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots).$$

This completes the proof of the lemma.

Define

$$O_n(\boldsymbol{\theta}) = \frac{1}{n} \log \frac{L_n(\boldsymbol{\theta})}{L_n(\boldsymbol{\theta}_0)}.$$

Lemma 4. For all $\boldsymbol{\theta} \in \Theta$, almost surely

$$\lim_{n\to\infty} O_n(\boldsymbol{\theta}) \leq 0$$

and the limit is almost surely equal to zero if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Proof. By Lemma 3 and Jensen's inequality, we have

$$\lim_{n\to\infty} O_n(\boldsymbol{\theta}) = E_{\theta_0} \log \frac{g_\theta(\boldsymbol{\epsilon}_t | \boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots)}{g_{\theta_0}(\boldsymbol{\epsilon}_t | \boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots)} \leq \log E_{\theta_0} \frac{g_\theta(\boldsymbol{\epsilon}_t | \boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots)}{g_{\theta_0}(\boldsymbol{\epsilon}_t | \boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots)} = 0.$$

By Assumption 1, this limit equals to zero if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. This completes the proof of the lemma.

Lemma 5. For any $\boldsymbol{\theta}_1 \in \Theta$, $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_0$, there exists a neighborhood $V(\boldsymbol{\theta}_1)$ of $\boldsymbol{\theta}_1$ such that

$$\limsup_{n\to\infty} \sup_{\theta\in V(\theta_1)} O_n(\boldsymbol{\theta}) < 0 \quad \text{a.s.}$$

Proof. Let $V_r(\boldsymbol{\theta}_1)$ be the open sphere with center $\boldsymbol{\theta}_1$ and radius $1/r$ and define

$$S_{2n}^r = \sup_{\theta\in V_r(\theta_1)} \left\| \prod_{t=2}^{n} M_\theta(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_{n-t+2}) \right\|.$$

Because this matrix norm is multiplicative, we have

$$S_{2,n+m}^r \leq S_{2n}^r \cdot S_{n+1,n+m}^r,$$

that is

$$\log S_{2,n+m}^r \leq \log S_{2n}^r + \log S_{n+1,n+m}^r$$

for any positive integers n, m and r. From the Kingman (1973) ergodic theorem for subadditive processes we can obtain

$$\lim_{n\to\infty} \frac{1}{n} \log S_{2n}^r = \lambda_r(\boldsymbol{\theta}_1) := \inf_{n>1} \frac{1}{n} E_{\theta_0} \log S_{2n}^r, \quad P_{\theta_0} - \text{a.s.}$$

Recall that

$$\lambda(\boldsymbol{\theta}_0) := \lim_{n\to\infty} \frac{1}{n} \log \left\| \prod_{t=2}^{n} M_{\theta_0}(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_{n-t+2}) \right\| = \inf_{n>1} \frac{1}{n} E_{\theta_0} \log \left\| \prod_{t=2}^{n} M_{\theta_0}(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_{n-t+2}) \right\|.$$

Thus, by Lemma 3 and 4, there exists $\varepsilon > 0$ and $n_\varepsilon \in \mathbb{N}$ such that

$$\frac{1}{n_\varepsilon} E_{\theta_0} \log \left\| \prod_{t=2}^{n_\varepsilon} M_{\theta_1}(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_{n-t+2}) \right\| < \lambda(\boldsymbol{\theta}_0) - \varepsilon.$$

The dominated convergence theorem shows that, for r large enough,

$$\lambda_r(\boldsymbol{\theta}_1) \leq \frac{1}{n_\varepsilon} E_{\theta_0} \log S_{2n_\varepsilon}^r < \lambda(\boldsymbol{\theta}_0) - \frac{\varepsilon}{2}.$$

This completes the proof of the lemma.

Proof of Theorem 3. Assume that $\hat{\boldsymbol{\theta}}_n$ didn't tend to $\boldsymbol{\theta}_0$ as $n \to \infty$, i.e., for arbitrarily large integer M, there exists a $\delta^* > 0$ and at least one $n^* \geq M$ such that $|\hat{\boldsymbol{\theta}}_{n^*} - \boldsymbol{\theta}_0| > \delta^*$ with positive probability. By Lemma 5, it follows that $L_{\hat{\theta}_{n^*}}(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_{n^*})$ is strictly less than $L_{\theta_0}(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_{n^*})$ with positive probability. However, with probability one, we have

$$L_{\hat{\theta}_n}(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n) = \sup_{\theta\in\Theta} L_\theta(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n) \geq L_{\theta_0}(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n)$$

for all n. The contradiction gives our result. This completes the proof of the theorem.

B Lagrange Multiplier (LM) test for constant within–regime correlations

The MS–GARCH model with constant within–regime correlations is attractive since it is analytically tractable. E.g., straightforward–to–check conditions for stationarity have been obtained, as well as a simple recursion for calculating multi–step conditional covariance matrices, which is crucial for mean–variance portfolio optimization. Moreover, in some simple cases, such as Pelletier's (2006) model with regime–independent GARCH dynamics, estimation in high dimensions is feasible via a two–step procedure with an embedded EM algorithm. Despite its convenience, it is still desirable to test whether the assumption of constant within–regime correlation matrices is tenable, since otherwise further improvement of out–of–sample portfolio selection might be feasible by extending the model to allow for within–regime correlation dynamics as in Billio and Caporin (2005) and Otranto (2010). Using results of Hamilton (1996), we extend the Lagrange Multiplier (LM) test devised by Tse (2000) for constant conditional correlations in multivariate GARCH models. In Tse (2000) the LM test is derived under normality of the innovations, but he reports simulations indicating it being quite robust against nonnormality. However, in view of the discussion in Section 4.1, this cannot be expected to hold for MS–GARCH processes, and thus we derive the test allowing for Student's t errors. The test under normality is then straightforwardly obtained if the degrees of freedom $\nu \to \infty$.

For the volatility dynamics, we assume that the conditional standard deviation of asset i in regime j, σ_{ijt}, is described by a standard (symmetric) AGARCH(1,1) process, i.e.,[15]

$$\sigma_{ijt} = \omega_{ij} + a_{ij}|\epsilon_{i,t-1}| + b_{ij}\sigma_{ij,t-1}, \quad i = 1, \dots, M, \quad j = 1, \dots, k. \tag{B.1}$$

The conditional correlation matrix in regime j is

$$\boldsymbol{R}_{jt} = (\rho_{i\ell,jt})_{i,\ell=1,\dots,M}, \quad j = 1, \dots, k.$$

where, as in Tse (2000), correlations evolve according to[16]

$$\rho_{i\ell,jt} = \overline{\rho}_{i\ell,j} + \delta_{i\ell,j}\epsilon_{i,t-1}\epsilon_{\ell,t-1}, \quad i = 1, \dots, M-1, \quad \ell = i+1, \dots, M, \quad j = 1, \dots, k. \tag{B.2}$$

[15] The extension to allow for asymmetric response of volatility is straightforward.

[16] As noted by Silvennoinen and Teräsvirta (2009b), (B.2) does not represent a specific alternative to the CCC as positive definite correlation matrices are not guaranteed for every t: "For this reason we interpret the test as a general misspecification test". Smith (2008) reports simulation results in favor of Hamilton's (1996) LM specification tests for Markov–switching models..

The null hypothesis of constant conditional within–regime correlations corresponds to

$$\mathrm{H}_0 : \delta_{i\ell,j} = 0, \qquad i = 1, \ldots, M-1, \quad \ell = i+1, \ldots, M, \quad j = 1, \ldots, k. \tag{B.3}$$

We distinguish between the following cases, which differ in the specification of the alternative hypothesis:

(a) The conditional correlation dynamics are unrestricted across regimes. In this case, under (B.3), the LM test statistic is asymptotically distributed as χ^2 with $kM(M-1)/2$ degrees of freedom.

(b) The conditional correlation dynamics are the same across regimes, i.e., $\delta_{i\ell,1} = \delta_{i\ell,2} = \cdots = \delta_{i\ell,k}$, $i = 1, \ldots, M-1$, $\ell = i+1, \ldots, M$. In this case, under (B.3), the LM test statistic is asymptotically distributed as χ^2 with $M(M-1)/2$ degrees of freedom.

For the purpose of the current section, it is convenient to decompose the parameter vector of the model as $\boldsymbol{\theta} = (\mathrm{vec}(\boldsymbol{P})', \boldsymbol{\vartheta}')'$,[17] where $\boldsymbol{\vartheta}$ consists of the parameters of the conditional regime densities, i.e., $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_1', \ldots, \boldsymbol{\vartheta}_k', \nu)'$, where ν is the (common) shape parameter of the t distribution and $\boldsymbol{\vartheta}_j = (\boldsymbol{\vartheta}_{1j}', \ldots, \boldsymbol{\vartheta}_{Mj}', \boldsymbol{\rho}_j', \boldsymbol{\delta}_j')'$, $j = 1, \ldots, k$, where $\boldsymbol{\vartheta}_{ij} = (\omega_{ij}, a_{ij}, b_{ij})'$, and $\boldsymbol{\rho}_j$ and $\boldsymbol{\delta}_j$ are the $M(M-1)/2$ vectors which stack, respectively, parameters $\overline{\rho}_{i\ell,j}$ and $\delta_{i\ell,j}$ in Equation (B.2), i.e.,

$$\begin{aligned}
\boldsymbol{\rho}_j &= (\overline{\rho}_{12,j}, \overline{\rho}_{13,j}, \ldots, \overline{\rho}_{1M,j}, \overline{\rho}_{23,j}, \ldots, \overline{\rho}_{2M,j}, \ldots, \overline{\rho}_{M-1,M,j})', \\
\boldsymbol{\delta}_j &= (\delta_{12,j}, \delta_{13,j}, \ldots, \delta_{1M,j}, \delta_{23,j}, \ldots, \delta_{2M,j}, \ldots, \delta_{M-1,M,j})'.
\end{aligned}$$

The log–likelihood of the model for a sample of size T is given by

$$\begin{aligned}
\log L(\boldsymbol{\theta}) &= \sum_{t=1}^{T} \log f(\epsilon_t | \Omega_{t-1}; \boldsymbol{\theta}) \\
&= \sum_{t=1}^{T} \log \left\{ \sum_{j=1}^{k} p(\Delta_t = j | \Omega_{t-1}; \boldsymbol{\theta}) f(\epsilon_t | \Omega_{t-1}, \Delta_t = j; \boldsymbol{\vartheta}_j, \nu) \right\},
\end{aligned} \tag{B.4}$$

where $\Omega_t = \{\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_{t-1}, \ldots, \boldsymbol{\epsilon}_0\}$, $p(\Delta_t = j | \Omega_{t-1}; \boldsymbol{\theta})$ are the one–step predicted regime inferences (cf. Hamilton, 1994, Ch. 22), and the conditional regime densities are

$$\begin{aligned}
f(\epsilon_t | \Omega_{t-1}, \Delta_t = j; \boldsymbol{\vartheta}_j, \nu) &= \frac{\Gamma\left(\frac{\nu+M}{2}\right)}{\Gamma(\nu/2)(\pi(\nu-2))^{M/2} |\boldsymbol{R}_{jt}|^{1/2} \prod_{i=1}^{M} \sigma_{ijt}} \left\{ 1 + \frac{d_{jt}^2}{\nu - 2} \right\}^{-(\nu+M)/2} \\
&=: f_{jt}(\boldsymbol{\vartheta}_j, \nu),
\end{aligned} \tag{B.5}$$

[17] Of course the last row of $\boldsymbol{P}$ is redundant.

where the squared Mahalanobis distance

$$d_{jt}^2 = \boldsymbol{\epsilon}_t' \boldsymbol{D}_{jt}^{-1} \boldsymbol{R}_{jt}^{-1} \boldsymbol{D}_{jt}^{-1} \boldsymbol{\epsilon}_t = \boldsymbol{\epsilon}_{jt}^{\star\prime} \boldsymbol{R}_{jt}^{-1} \boldsymbol{\epsilon}_{jt}^{\star} = \boldsymbol{\epsilon}_{jt}^{\star\prime} \tilde{\boldsymbol{\epsilon}}_{jt}, \tag{B.6}$$

with

$$\boldsymbol{\epsilon}_{jt}^{\star} = \boldsymbol{D}_{jt}^{-1} \boldsymbol{\epsilon}_t, \quad \tilde{\boldsymbol{\epsilon}}_{jt} = \boldsymbol{R}_{jt}^{-1} \boldsymbol{\epsilon}_{jt}^{\star}, \quad j = 1, \ldots, k. \tag{B.7}$$

The regime–specific log–density for observation t is

$$\begin{aligned} \log f_{jt}(\boldsymbol{\vartheta}_j, \nu) &= -\frac{M}{2}(\log \pi + \log(\nu - 2)) + \log \Gamma\left(\frac{\nu + M}{2}\right) - \log \Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2} \log |\boldsymbol{R}_{jt}| \\ &\quad - \sum_{i=1}^{M} \log \sigma_{ijt} + \frac{\nu + M}{2} \log\left(1 + \frac{d_{jt}^2}{\nu - 2}\right) \\ &= -\frac{M}{2} \log \pi + \log \Gamma\left(\frac{\nu + M}{2}\right) - \log \Gamma\left(\frac{\nu}{2}\right) + \frac{\nu}{2} \log(\nu - 2) - \frac{1}{2} \log |\boldsymbol{R}_{jt}| \\ &\quad - \sum_{i=1}^{M} \log \sigma_{ijt} - \frac{\nu + M}{2} \log(\nu - 2 + d_{jt}^2), \quad j = 1, \ldots, k. \end{aligned} \tag{B.8}$$

The partial derivatives of (B.8) are obtained as

$$\begin{aligned} \frac{\partial \log f_{jt}(\boldsymbol{\vartheta}_j, \nu)}{\partial \nu} &= \frac{1}{2}\left[\psi\left(\frac{\nu + M}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right] - \frac{1}{2} \log\left(1 + \frac{d_{jt}^2}{\nu - 2}\right) \\ &\quad + \frac{1}{2}\left[\frac{\nu}{\nu - 2} - \frac{\nu + M}{\nu - 2 + d_{jt}^2}\right], \end{aligned} \tag{B.9}$$

$$\frac{\partial \log f_{jt}(\boldsymbol{\vartheta}_j, \nu)}{\partial \boldsymbol{\vartheta}_{ij}} = \frac{\partial \sigma_{ijt}}{\partial \boldsymbol{\nu}_{ij}} \frac{1}{\sigma_{ijt}} \left(\frac{\epsilon_{ijt}^{\star} \tilde{\epsilon}_{ijt}(\nu + M)}{\nu - 2 + d_{jt}^2} - 1\right), \quad i = 1, \ldots, M, \tag{B.10}$$

$$\frac{\partial \log f_{jt}(\boldsymbol{\vartheta}_j, \nu)}{\partial \boldsymbol{\rho}_j} = \frac{1}{2} \boldsymbol{U} \left(\frac{\nu + M}{\nu - 2 + d_{jt}^2} (\tilde{\boldsymbol{\epsilon}}_{jt} \otimes \tilde{\boldsymbol{\epsilon}}_{jt}) - \mathrm{vec} \boldsymbol{R}_{jt}^{-1}\right), \tag{B.11}$$

$$\frac{\partial \log f_{jt}(\boldsymbol{\vartheta}_j, \nu)}{\partial \boldsymbol{\delta}_j} = \frac{1}{2} \boldsymbol{U} \left\{\left(\frac{\nu + M}{\nu - 2 + d_{jt}^2} (\tilde{\boldsymbol{\epsilon}}_{jt} \otimes \tilde{\boldsymbol{\epsilon}}_{jt}) - \mathrm{vec} \boldsymbol{R}_{jt}^{-1}\right) \odot (\boldsymbol{\epsilon}_{t-1} \otimes \boldsymbol{\epsilon}_{t-1})\right\}, \tag{B.12}$$

where $\psi(x) = \mathrm{d} \log \Gamma(x)/\mathrm{d}x$ is the logarithmic derivative of the gamma function (the digamma function); $\epsilon_{ijt}^{\star}$ and $\tilde{\epsilon}_{ijt}$ are, respectively, the ith elements of vectors $\boldsymbol{\epsilon}_{jt}^{\star}$ and $\tilde{\boldsymbol{\epsilon}}_{jt}$ defined in (B.7) $i = 1, \ldots, M$; and the $M(M-1)/2 \times M^2$ matrix $\boldsymbol{U}$ is defined as in Silvennoinen and Teräsvirta (2009a), i.e., with its $[(i-1)M - i(i+1)/2 + \ell]$th row given by

$$\mathrm{vec}(\tilde{\boldsymbol{e}}_i \tilde{\boldsymbol{e}}_\ell' + \tilde{\boldsymbol{e}}_\ell \tilde{\boldsymbol{e}}_i')', \quad i = 1, \ldots, M-1, \quad \ell = i+1, \ldots, M,$$

where $\tilde{\boldsymbol{e}}_i$ is the ith column of the M–dimensional identity matrix. The derivative in (B.10) is

$$\frac{\partial \sigma_{ijt}}{\partial \boldsymbol{\vartheta}_{ij}} = \boldsymbol{\eta}_{ij,t-1} + b_{ij} \frac{\partial \sigma_{ij,t-1}}{\partial \boldsymbol{\vartheta}_{ij}}, \quad t = 2, \ldots, T, \tag{B.13}$$

where $\boldsymbol{\eta}_{ijt} = (1, |\epsilon_{it}|, \sigma_{ijt})'$, and the starting value in recursion (B.13) is

$$\frac{\partial \sigma_{ij,t=1}}{\partial \boldsymbol{\vartheta}_{ij}} = (1, |\epsilon_{i0}|, \sigma_{ij0})', \tag{B.14}$$

where we initialize all regime–specific conditional standard deviations with the sample standard deviation, i.e., in (B.14),

$$\sigma_{ij0} = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}\epsilon_{it}^2}, \quad i = 1, \ldots, M, \quad j = 1, \ldots, k.$$

The score of the tth observation is given by the derivative of the conditional log–density of $\boldsymbol{\epsilon}_t$ as given in (B.4) and (B.5),

$$\frac{\partial \log f(\epsilon_t|\Omega_{t-1};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial \log \left\{\sum_{j=1}^{k} p(\Delta_t = j|\Omega_{t-1};\boldsymbol{\theta}) f_{jt}(\boldsymbol{\vartheta}_j, \nu)\right\}}{\partial \boldsymbol{\theta}}. \tag{B.15}$$

Hamilton (1996) has shown that the derivatives in (B.15) involving elements of $\boldsymbol{\vartheta}$ can be evaluated as

$$\begin{aligned}
\frac{\partial \log f(\epsilon_t|\Omega_{t-1};\boldsymbol{\theta})}{\partial \boldsymbol{\vartheta}} &= \sum_{j=1}^{k} \frac{\partial \log f_{jt}(\boldsymbol{\vartheta}_j, \nu)}{\partial \boldsymbol{\vartheta}} p(\Delta_t = j|\Omega_t;\boldsymbol{\theta}) \\
&\quad + \sum_{\tau=1}^{t-1}\sum_{j=1}^{k} \frac{\partial f_{j\tau}(\boldsymbol{\vartheta}_j, \nu)}{\partial \boldsymbol{\vartheta}} \left[p(\Delta_\tau = j|\Omega_t;\boldsymbol{\theta}) - p(\Delta_\tau = j|\Omega_{t-1};\boldsymbol{\theta})\right], \\
t &= 1, \ldots, T,
\end{aligned} \tag{B.16}$$

where the second line in (B.16) is set to zero for $t = 1$. For $\tau = t$ and $\tau < t$, the regime probabilities $p(\Delta_\tau = j|\Omega_t;\boldsymbol{\theta})$ in (B.16) are known as *filtered* and *smoothed* regime inferences, respectively; see, e.g., Hamilton (1994, Ch. 22) and Kim (1994) for their recursive calculation. We initialize these recursions by assuming that Δ_1 is drawn from the stationary distribution of the chain, i.e.,

$$\begin{aligned}
p(\Delta_1 = 1|\boldsymbol{\theta}) &= \pi_{1,\infty} = \frac{1 - p_{22}}{2 - p_{11} - p_{22}}, \\
p(\Delta_1 = 2|\boldsymbol{\theta}) &= \pi_{2,\infty} = \frac{1 - p_{11}}{2 - p_{11} - p_{22}}.
\end{aligned} \tag{B.17}$$

Note that, in (B.16), many terms are zero for parameters that appear in only one regime; e.g., if all parameters except ν are regime–specific, then

$$\begin{aligned}
\frac{\partial \log f(\epsilon_t|\Omega_{t-1};\boldsymbol{\theta})}{\partial \boldsymbol{\vartheta}_j} &= \frac{\partial \log f_{jt}(\boldsymbol{\vartheta}_j, \nu)}{\partial \boldsymbol{\vartheta}_j} p(\Delta_t = j|\Omega_t;\boldsymbol{\theta}) \\
&\quad + \sum_{\tau=1}^{t-1} \frac{\partial f_{j\tau}(\boldsymbol{\vartheta}_j, \nu)}{\partial \boldsymbol{\vartheta}_j} \left[p(\Delta_\tau = j|\Omega_t;\boldsymbol{\theta}) - p(\Delta_\tau = j|\Omega_{t-1};\boldsymbol{\theta})\right] \\
t &= 1, \ldots, T, \quad j = 1, \ldots, k.
\end{aligned}$$

For the score with respect to the parameters of $\boldsymbol{P}$, see Hamilton (1996).[18]

Now let $\boldsymbol{S}$ be the $T \times N$ matrix (where N is the dimension of $\boldsymbol{\theta}$) the tth row of which is given by (the transpose of) the score (B.15), $t = 1, \ldots, T$, and let $\widehat{\boldsymbol{S}}$ be $\boldsymbol{S}$ evaluated at $\widehat{\boldsymbol{\theta}}^c$ the constrained MLE of $\boldsymbol{\theta}$ under (B.3). Then the LM test statistic for H_0 given by (B.3) can be calculated as (in the *outer gradient product* (OPG) form, cf. Godfrey, 1988, p. 15; and Hamilton, 1996)

$$\mathrm{LM} = \mathbf{1}_T' \widehat{\boldsymbol{S}} (\widehat{\boldsymbol{S}}' \widehat{\boldsymbol{S}})^{-1} \widehat{\boldsymbol{S}}' \mathbf{1}_T \xrightarrow{d} \chi^2(c_0), \tag{B.18}$$

where $\mathbf{1}_T$ is a T–dimensional column of ones, and c_0 is the number of parameter constraints under H_0, see the discussion following Equation (B.3). Note the elements of $\mathbf{1}_T'\widehat{\boldsymbol{S}}$ corresponding to unrestricted parameters are zero, so that a slight simplification of the LM statistic (B.18) can be obtained (cf. Engle, 1984, p. 783).

References

Abramson, A. and Cohen, I. (2007). On the Stationarity of Markov–switching GARCH Processes. *Econometric Theory*, 23:485–500.

Alexander, C. (2008). *Practical Financial Econometrics.* John Wiley & Sons, Chichester.

Ang, A. and Timmermann, A. (2012). Regime Changes and Financial Markets. *Annual Review of Financial Economics*, 4:313–337.

Ardia, D. (2009). Bayesian Estimation of a Markov–switching Threshold Asymmetric GARCH Model with Student-t Innovations. *Econometrics Journal*, 12:105–126.

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

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

Bollerslev, T. (1990). Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model. *Review of Economics and Statistics*, 73:498–505.

Bougerol, P. and Picard, N. (1992). Strict stationarity of generalized autoregressive process. *Annals of Probability*, 20:1714–1730.

Brandt, A. (1986). The stochastic equation with stationary coefficients. *Adv. Apl. Probab.*, 18:221–254.

Broda, S. A., Haas, M., Krause, J., Paolella, M. S., and Steude, S. C. (2013). Stable Mixture GARCH Models. *Journal of Econometrics*, 172:292–306.

Bulla, J. (2011). Hidden Markov models with t components. Increased persistence and other aspects. *Quantitative Finance*, 11:459–475.

[18] In particular, for $k = 2$ regimes and initial probabilities (B.17), the relevant derivatives with respect to p_{11} and p_{22} are given by Equations (3.15) and (3.16) in Hamilton (1996).

Campbell, R., Koedijk, K., and Kofman, P. (2002). Increased Correlation in Bear Markets: A Downside Risk Perspective. *Financial Analysts Journal*, 58:87–94.

Case, B., Guidolin, M., and Yildrim, Y. (2014). Markov Switching Dynamics in REIT Returns: Univariate and Multivariate Evidence on Forecasting Performance. *Real Estate Economics*, 42:279–342.

Conrad, C. and Karanasos, M. (2010). Negative volatility spillovers in the unrestricted ECCC-GARCH model. *Econometric Theory*, 26:838–862.

Dean, W. G. and Faff, R. W. (2008). Evidence of Feedback Trading with Markov Switching Regimes. *Review of Quantitative Finance and Accounting*, 30:133–151.

Ding, Z., Granger, C. W. J., and Engle, R. F. (1993). A Long Memory Property of Stock Market Returns and a New Model. *Journal of Empirical Finance*, 1:83–106.

Dueker, M. J. (1997). Markov Switching in GARCH Processes and Mean–Reverting Stock–Market Volatility. *Journal of Business and Economic Statistics*, 15:26–34.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50:987–1008.

Engle, R. F. (1984). Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics. In Griliches, Z. and Intriligator, M. D., editors, *Handbook of Econometrics*, volume 2, chapter 13, pages 775–826. Elsevier, Amsterdam.

Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics*, 20:339–350.

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

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

Francq, C. and Zakoïan, J.-M. (2008). Deriving the Autocovariances of Powers of Markov–switching GARCH Models, with Applications to Statistical Inference. *Computational Statistics & Data Analysis*, 52:3027–3046.

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

Giacomini, R., Gottschling, A., Haefke, C., and White, H. (2008). Mixtures of t-distributions for finance and forecasting. *Journal of Econometrics*, 144:175–192.

Giot, P. and Laurent, S. (2003). Value–at–Risk for Long and Short Trading Positions. *Journal of Applied Econometrics*, 18:641–664.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance*, 48:1779–1801.

Godfrey, L. G. (1988). *Misspecification Tests in Econometrics: The Lagrange Multiplier Principle and Other Approaches.* Cambridge University Press, Cambridge.

Gray, S. F. (1996). Modeling the Conditional Distribution of Interest Rates as a Regime–Switching Process. *Journal of Financial Economics*, 42:27–62.

Guidolin, M. (2011). Markov Switching Models in Empirical Finance. In Drukker, D. M., editor, *Missing Data Methods: Time-Series Methods and Applications (Advances in Econometrics, Volume 27)*, pages 1–86. Emerald, Bingley.

Haas, M. (2009). Value-at-Risk via Mixture Distributions Reconsidered. *Applied Mathematics and Computation*, 215:2103–2119.

Haas, M. (2010). Covariance Forecasts and Long-run Correlations in a Markov-switching Model for Dynamic Correlations. *Finance Research Letters*, 7:86–97.

Haas, M., Mittnik, S., and Paolella, M. S. (2004). A New Approach to Markov–Switching GARCH Models. *Journal of Financial Econometrics*, 2:493–530.

Haas, M. and Paolella, M. S. (2012). Mixture and Regime-switching GARCH Models. In Bauwens, L., Hafner, C. M., and Laurent, S., editors, *Handbook of Volatility Models and their Applications*.

Hamilton, J. D. (1996). Specification Testing in Markov–Switching Time–Series Models. *Journal of Econometrics*, 70:127–157.

He, C. and Teräsvirta, T. (2004). An Extended Constant Correlation GARCH Model and its Fourth-Moment Structure. *Econometric Theory*, 20:904–926.

Heaney, R. and Sriananthakumar, S. (2012). Time-varying correlation between stock market returns and real estate returns. *Journal of Empirical Finance*, 19:583594.

Jeantheau, T. (1998). Strong Consistency of Estimators for Multivariate ARCH Models. *Econometric Theory*, 14:70–86.

Kasch, M. and Caporin, M. (2013). Volatility Threshold Dynamic Conditional Correlations: An International Analysis. *Journal of Financial Econometrics*, 11:706–742.

Keribin, C. (2000). Consistent estimation of the order of mixture models. *Sankhyā: The Indian Journal of Statistics, Series A*, 62:49–66.

Kingman, J. (1973). Subadditive ergodic theory. *Annals of Probability*, 1:883–899.

Klaassen, F. (2002). Improving GARCH Volatility Forecasts with Regime–Switching GARCH. *Empirical Economics*, 27:363–394.

Ledoit, O., Santa-Clara, P., and Wolf, M. (2003). Flexible Multivariate GARCH Modeling with an Application to International Stock Markets. *Review of Economics and Statistics*, 85:735–747.

Lee, Y.-H. (2014). An international analysis of REITs and stock portfolio management based on dynamic conditional correlation models. *Financial Markets and Portfolio Management*, 28:165–180.

Lejeune, B. (2009). A Diagnostic M-test for Distributional Specification of Parametric Conditional Heteroscedasticity Models for Financial Data. *Journal of Empirical Finance*, 16:507–523.

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

Liu, J.-C. (2007). Stationarity for a Markov–switching Box–Cox Transformed Threshold GARCH Process. *Statistics and Probability Letters*, 77:1428–1438.

McAleer, M., Hoti, S., and Chan, F. (2009). Structure and Asymptotic Theory for Multivariate Asymmetric Conditional Volatility. *Econometric Reviews*, 28:422–440.

Mittnik, S. (2014). VaR-implied tail-correlation matrices. *Economics Letters*, 122:69–73.

Morawski, J., Rehkugler, H., and Füss, R. (2008). The nature of listed real estate companies: property or equity market? *Financial Markets and Portfolio Management*, 22:101–126.

Nabeya, S. (1951). Absolute Moments in 2-dimensional Normal Distribution. *Annals of the Institute of Statistical Mathematics*, 3:2–6.

Nakatani, T. and Teräsvirta, T. (2008). Positivity constraints on the conditional variances in the family of conditional correlation GARCH models. *Finance Research Letters*, 5:88–95.

Nakatani, T. and Teräsvirta, T. (2009). Testing for volatility interactions in the Constant Conditional Correlation GARCH model. *Econometrics Journal*, 12:147–163.

Nikolaev, N. Y., Boshnakov, G. N., and Zimmer, R. (2013). Heavy-tailed mixture GARCH volatility modeling and Value-at-Risk estimation. *Expert Systems with Applications*, 40:2233–2243.

Otranto, E. (2010). Asset Allocation using Flexible Dynamic Correlation Models with Regime Switching. *Quantitative Finance*, 10:325–338.

Pelletier, D. (2006). Regime Switching for Dynamic Correlations. *Journal of Econometrics*, 131:445–473.

Ramchand, L. and Susmel, R. (1998). Volatility and Cross Correlation Across Major Stock Markets. *Journal of Empirical Finance*, 5:397–416.

Ryden, T., Teräsvirta, T., and Åsbrink, S. (1998). Stylized Facts of Daily Return Series and the Hidden Markov Model. *Journal of Applied Econometrics*, 13:217–244.

Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics*, 6:461–464.

Seneta, E. (1981). *Nonnegative Matrices and Markov Chains*. Springer, New York.

Silvennoinen, A. and Teräsvirta, T. (2009a). Modeling Multivariate Autoregressive Conditional Heteroskedasticity with the Double Smooth Transition Conditional Correlation GARCH Model. *Journal of Financial Econometrics*, 7:373–411.

Silvennoinen, A. and Teräsvirta, T. (2009b). Multivariate GARCH Models. In Mikosch, T., Krei, J.-P., Davis, R. A., and Andersen, T. G., editors, *Handbook of Financial Time Series*, pages 201–229. Springer, Berlin.

Smith, D. R. (2008). Evaluating Specification Tests for Markov-Switching Time-Series Models. *Journal of Time Series Analysis*, 29:629–652.

Taylor, S. J. (1986). *Modelling Financial Time Series*. John Wiley & Sons, Chichester.

Tse, Y. K. (2000). A test for constant correlations in a multivariate GARCH model. *Journal of Econometrics*, 98:107–127.

Wong, C. S., Chan, W. S., and Ka, P. L. (2009). A Student t-mixture autoregressive model with applications to heavy-tailed financial data. *Biometrika*, 96:751–760.