SUPPLEMENTARY MATERIAL for
"HIGHER ORDER MOMENTS OF
MARKOV SWITCHING VARMA MODELS"

Maddalena Cavicchioli

1. A review on Francq and Zakoïan (2001)'s results

The following theorem, proved in Francq and Zakoïan (2001), FZ01, gives conditions for the existence of second order stationary Markov switching MS VARMA processes.

Result 1.1 (FZ01, Theorem 2, p.346) *Suppose that* $\rho(\mathbf{P}(\boldsymbol{\Phi}^{\otimes 2})) < 1$, *where* $\rho(\cdot)$ *denotes the spectral radius. Then, for all* $t \in \mathbb{Z}$, *the series*

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

converges in L^2 *and the* MS(M) VARMA(p,q) *process* $(\mathbf{y}_t)$, *defined as the block of the first* K *components of* $(\mathbf{z}_t)$, *is the unique nonanticipative second order stationary solution of* (1). *Suppose that* (2) *admits a nonanticipative second order stationary solution. Then we have*

$$\sum_{k=0}^{+\infty} ||\mathbf{I}\,\{\mathbf{P}(\boldsymbol{\Phi}^{\otimes 2})\}^k \, \mathbf{S} \; \mathrm{vec}\, \boldsymbol{\Omega}|| < +\infty$$

where $\mathbf{I} = (\mathbf{I}_{n^2} \cdots \mathbf{I}_{n^2}) \in \mathbb{R}^{n^2 \times (Mn^2)}$ *and* $||\cdot||$ *denotes the matrix norm* $||\mathbf{A}|| = \sum_{i,j} |a_{ij}|$ *for any matrix* $\mathbf{A} = (a_{ij})$. *Finally, if* $\mathbf{c}(s_t) = \mathbf{0}$ *in* (1), *a necessary and sufficient condition for the existence of a nonanticipative second order stationary solution to* (1) *is given by the finiteness of the above series.*

Once first and second order stationarity are ensured, FZ01, Section 4, compute the mean and the variance-covariance matrix of the process $\mathbf{y} = (\mathbf{y}_t)$ in (1). Let $\mathbf{U}$ be the (Mn)-dimensional vector, whose ith block is the n-dimensional vector $\pi_i \, E(\mathbf{z}_t \,|\, s_t = i)$ for $i = 1, \ldots, M$, that is,

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

Let $\mathbf{c} = (\pi_1 \, \mathbf{c}(1)^{'} \quad \cdots \quad \pi_M \, \mathbf{c}(M)^{'})^{'} \in \mathbb{R}^{Mn}$. Then we have

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

(see FZ01, p.348). Costa *et al.* (2005), Proposition 3.6, p.35, proved that if $\rho(\mathbf{P}(\mathbf{\Phi}^{\otimes 2})) < 1$, then $\rho(\mathbf{P}(\mathbf{\Phi})) < 1$. It can be easily checked that the converse is not true in general. See Remark 3.7, p.35, of Costa *et al.* (2005). We extend in Section 3, Theorem 3.3, such a result for the general case of a matrix $\mathbf{P}(\mathbf{\Phi}^{\otimes r})$.

Result 1.2 (FZ01, Section 4.1) *If* $\rho(\mathbf{P}(\mathbf{\Phi})) < 1$*, then the expectations of* $(\mathbf{z}_t)$ *in* (2) *and* $(\mathbf{y}_t)$ *in* (1)*, driven by a* MS(M) VARMA(p, q) *model, are given by*

$$E(\mathbf{z}_t) = (\mathbf{e}^{'} \otimes \mathbf{I}_n) \, \mathbf{U} \qquad\qquad E(\mathbf{y}_t) = (\mathbf{e}^{'} \otimes \mathbf{f}^{'}) \, \mathbf{U}$$

where

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

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

Let $\mathbf{V}$ be defined as $\mathbf{U}$ but with $\mathbf{z}_t$ replaced by $\mathbf{z}_t \otimes \mathbf{z}_t = \text{vec}(\mathbf{z}_t \, \mathbf{z}_t')$, that is,

$$\mathbf{V} = (\pi_1 \, E(\mathbf{z}_t' \otimes \mathbf{z}_t' | \, s_t = 1) \quad \cdots \quad \pi_M \, E(\mathbf{z}_t' \otimes \mathbf{z}_t' | \, s_t = M))' \in \mathbb{R}^{Mn^2}.$$

Let $\mathbf{D}$ be defined by replacing $\mathbf{\Phi}(i) \otimes \mathbf{\Phi}(i)$ by $\mathbf{c}(i) \otimes \mathbf{\Phi}(i) + \mathbf{\Phi}(i) \otimes \mathbf{c}(i)$ in the definition of $\mathbf{P}(\mathbf{\Phi}^{\otimes 2})$. Then we have (see FZ01, p.349)

$$\mathbf{V} = \mathbf{P}(\mathbf{\Phi}^{\otimes 2}) \, \mathbf{V} \, + \, \mathbf{D} \, \mathbf{U} \, + \, \mathbf{C} \, + \, \mathbf{S} \, \text{vec}(\mathbf{\Omega}).$$

Result 1.3 (FZ01, Section 4.1) *If* $\rho(\mathbf{P}(\mathbf{\Phi}^{\otimes 2})) < 1$, *then the second order moments of* $(\mathbf{z}_t)$ *in* (2) *and* $(\mathbf{y}_t)$ *in* (1), *driven by a* MS(M) VARMA(p, q) *model, are given by*

$$E(\mathbf{z}_t \otimes \mathbf{z}_t) = (\mathbf{e}' \otimes \mathbf{I}_{n^2}) \, \mathbf{V} \qquad\qquad E(\mathbf{y}_t \otimes \mathbf{y}_t) = (\mathbf{e}' \otimes \mathbf{f}' \otimes \mathbf{f}') \, \mathbf{V}$$

where

$$\mathbf{V} = (\mathbf{I}_{Mn^2} \, - \, \mathbf{P}(\mathbf{\Phi}^{\otimes 2}))^{-1} \, [\mathbf{D} \, \mathbf{U} \, + \, \mathbf{C} \, + \, \mathbf{S} \, \text{vec}(\mathbf{\Omega})].$$

DGP	Moment	$T = 100$	$T = 250$	$T = 500$	$T = 1000$	$T = 2000$	$T = 5000$
Bivariate MS(2) VAR(1)	1st	0.1083	0.0661	0.0487	0.0329	0.0246	0.0152
	2nd	0.1798	0.1610	0.1563	0.1501	0.1483	0.0864
	3rd	0.2342	0.1597	0.1260	0.0897	0.0696	0.0462
	4th	0.6014	0.4909	0.4439	0.3867	0.3587	0.1413
Bivariate MS(2) VARMA(1,1)	1st	0.1747	0.1051	0.0772	0.0532	0.0389	0.0244
	2nd	0.4602	0.4105	0.3947	0.3784	0.3735	0.1686
	3rd	0.8605	0.5860	0.4706	0.3311	0.2571	0.0571
	4th	3.4824	2.9544	2.6749	2.3861	2.2109	0.9826
Univariate MS(2) ARMA(2,1)	1st	0.0958	0.0599	0.0403	0.0299	0.0209	0.0138
	2nd	0.1406	0.1061	0.0892	0.0774	0.0737	0.0210
	3rd	0.3026	0.2148	0.1598	0.1316	0.0898	0.0350
	4th	1.2468	1.0083	0.8375	0.7237	0.6007	0.2741

Table 1: RMSE of the discrepancies between analytical and empirical first, second, third and fourth moments (vertically ordered) over 1,000 replications. The DGPs and their parameters are described above.

2. Monte Carlo results

To check the correctness of the formulae proposed in Section 3, we run some Monte Carlo experiments. Particularly, we aim at checking if the proposed analytical moments approach their empirical counterparts. To do so, we simulate time series from three different data generating processes: a bivariate ($K = 2$) MS(2) VAR(1), a bivariate ($K = 2$) MS(2) VARMA$(1, 1)$, and a univariate ($K = 1$) MS(2) ARMA$(2, 1)$. The coefficients of the simulated bivariate MS(2) VAR(1) are as follows (consider the model in Equation (1) with no constant term and Gaussian i.i.d. errors): $p_{11} = 0.1$, $p_{22} = 0.8$, $a_1(1) = \begin{pmatrix} 0 & 0.4 \\ 0.3 & 1.2 \end{pmatrix}$, $a_1(2) = \begin{pmatrix} 0.6 & 0.4 \\ 0 & 0.3 \end{pmatrix}$, $\sigma(1) = \begin{pmatrix} 1 & 0.6 \\ 0.6 & 1 \end{pmatrix}$ and $\sigma(2) =$

$\begin{pmatrix} 0.4 & 0.2 \\ 0.2 & 0.4 \end{pmatrix}$. In the case of the bivariate MS(2) VARMA(1,1), we add the moving average part, where $b_1(1) = \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0 \end{pmatrix}$ and $b_1(2) = \begin{pmatrix} 0 & 0.2 \\ 0.3 & 0.8 \end{pmatrix}$. Finally, for the univariate MS(2) ARMA(2,1), we set the following parameters: $a_1(1) = 1.2$, $a_1(2) = 0$, $a_2(1) = 0.1$, $a_2(2) = 0.3$, $b_1(1) = 0.5$, $b_1(2) = 0.5$, $\sigma(1) = 1$ and $\sigma(2) = 0.5$. The experiments simulate artificial time series of length $T + 50$ with $T = \{100, 250, 500, 1000, 2000, 5000\}$; the first 50 initial data points are discarded to minimize the effect of initial conditions. One thousand Monte Carlo replications are carried out for each trial. In Table 1 we report the RMSE over the simulations evaluated as the squared-norm of the difference between the empirical and the true moment vectors (vertically ordered from the first to the fourth moment). The numerical exercises show that our measures of the third and fourth moments, as well as first and second moments in FZ01, are very close to their empirical counterparts.

3. Proof of Theorem 3.2

First we prove Equation (4). Starting from (2), similar computations as in the proof of Theorem 3.1 show that

$$\begin{aligned}
&\pi_i\, E(\mathbf{z}_t^{\otimes 4} \,|\, s_t = i) = \pi_i\, (\mathbf{\Phi}^{\otimes 4}(i))\, E(\mathbf{z}_{t-1}^{\otimes 4} \,|\, s_t = i) + \pi_i\, [\mathbf{c}(i) \otimes \mathbf{\Phi}^{\otimes 3}(i) \\
&+ \mathbf{\Phi}(i) \otimes \mathbf{c}(i) \otimes \mathbf{\Phi}^{\otimes 2}(i) + \mathbf{\Phi}^{\otimes 2}(i) \otimes \mathbf{c}(i) \otimes \mathbf{\Phi}(i) + \mathbf{\Phi}^{\otimes 3}(i) \otimes \mathbf{c}(i)]\ E(\mathbf{z}_{t-1}^{\otimes 3} \,|\, s_t = i) \\
&+ \pi_i\, [\mathbf{\Phi}^{\otimes 2}(i) \otimes \mathbf{c}^{\otimes 2}(i) + (\mathbf{\Phi}(i) \otimes \mathbf{c}(i))^{\otimes 2} + \mathbf{\Phi}(i) \otimes \mathbf{c}^{\otimes 2}(i) \otimes \mathbf{\Phi}(i) \\
&+ \mathbf{c}(i) \otimes \mathbf{\Phi}^{\otimes 2}(i) \otimes \mathbf{c}(i) + (\mathbf{c}(i) \otimes \mathbf{\Phi}(i))^{\otimes 2} + \mathbf{c}^{\otimes 2}(i) \otimes \mathbf{\Phi}^{\otimes 2}(i) \\
&+ (\mathbf{\Phi}^{\otimes 2}(i) \otimes \mathbf{\Sigma}^{\otimes 2}(i))\, E(\mathbf{I}_n^{\otimes 2} \otimes \mathbf{u}_t^{\otimes 2}) + (\mathbf{\Phi}(i) \otimes \mathbf{\Sigma}(i))^{\otimes 2}\, E((\mathbf{I}_n \otimes \mathbf{u}_t)^{\otimes 2})
\end{aligned}$$

$$
\begin{aligned}
&+ (\boldsymbol{\Phi}(i) \otimes \boldsymbol{\Sigma}^{\otimes 2}(i) \otimes \boldsymbol{\Phi}(i))\, E(\mathbf{I}_n \otimes \mathbf{u}_t^{\otimes 2} \otimes \mathbf{I}_n) + (\boldsymbol{\Sigma}(i) \otimes \boldsymbol{\Phi}^{\otimes 2}(i) \otimes \boldsymbol{\Sigma}(i))\, E(\mathbf{u}_t \otimes \mathbf{I}_n^{\otimes 2} \otimes \mathbf{u}_t) \\
&+ (\boldsymbol{\Sigma}(i) \otimes \boldsymbol{\Phi}(i))^{\otimes 2}\, E((\mathbf{u}_t \otimes \mathbf{I}_n)^{\otimes 2}) + (\boldsymbol{\Sigma}^{\otimes 2}(i) \otimes \boldsymbol{\Phi}^{\otimes 2}(i))\, E(\mathbf{u}_t^{\otimes 2} \otimes \mathbf{I}_n^{\otimes 2})]\, E(\mathbf{z}_{t-1}^{\otimes 2} \,|\, s_t = i) \\
&+ \pi_i\, [\boldsymbol{\Phi}(i) \otimes \mathbf{c}^{\otimes 3}(i) + \mathbf{c}(i) \otimes \boldsymbol{\Phi}(i) \otimes \mathbf{c}^{\otimes 2}(i) + \mathbf{c}^{\otimes 2}(i) \otimes \boldsymbol{\Phi}(i) \otimes \mathbf{c}(i) \\
&+ \mathbf{c}^{\otimes 3}(i) \otimes \boldsymbol{\Phi}(i) + (\boldsymbol{\Phi}(i) \otimes \mathbf{c}(i) \otimes \boldsymbol{\Sigma}^{\otimes 2}(i))\, E(\mathbf{I}_n \otimes \mathbf{u}_t^{\otimes 2}) \\
&+ (\boldsymbol{\Phi}(i) \otimes \boldsymbol{\Sigma}(i) \otimes \mathbf{c}(i) \otimes \boldsymbol{\Sigma}(i))\, E(\mathbf{I}_n \otimes \mathbf{u}_t^{\otimes 2}) + (\boldsymbol{\Phi}(i) \otimes \boldsymbol{\Sigma}^{\otimes 2}(i) \otimes \mathbf{c}(i))\, E(\mathbf{I}_n \otimes \mathbf{u}_t^{\otimes 2}) \\
&+ (\mathbf{c}(i) \otimes \boldsymbol{\Phi}(i) \otimes \boldsymbol{\Sigma}^{\otimes 2}(i))\, E(\mathbf{I}_n \otimes \mathbf{u}_t^{\otimes 2}) + (\mathbf{c}(i) \otimes \boldsymbol{\Sigma}(i) \otimes \boldsymbol{\Phi}(i) \otimes \boldsymbol{\Sigma}(i)) \\
&\times\ E(\mathbf{u}_t \otimes \mathbf{I}_n \otimes \mathbf{u}_t) + (\mathbf{c}(i) \otimes \boldsymbol{\Sigma}^{\otimes 2}(i) \otimes \boldsymbol{\Phi}(i))\, E(\mathbf{u}_t^{\otimes 2} \otimes \mathbf{I}_n) + (\boldsymbol{\Sigma}(i) \otimes \boldsymbol{\Phi}(i) \otimes \mathbf{c}(i) \otimes \boldsymbol{\Sigma}(i)) \\
&\times\ E(\mathbf{u}_t \otimes \mathbf{I}_n \otimes \mathbf{u}_t) + (\boldsymbol{\Sigma}(i) \otimes \boldsymbol{\Phi}(i) \otimes \boldsymbol{\Sigma}(i) \otimes \mathbf{c}(i))\, E(\mathbf{u}_t \otimes \mathbf{I}_n \otimes \mathbf{u}_t) + (\boldsymbol{\Sigma}(i) \otimes \mathbf{c}(i) \\
&\otimes \boldsymbol{\Phi}(i) \otimes \boldsymbol{\Sigma}(i))\, E(\mathbf{u}_t \otimes \mathbf{I}_n \otimes \mathbf{u}_t) + (\boldsymbol{\Sigma}(i) \otimes \mathbf{c}(i) \otimes \boldsymbol{\Sigma}(i) \otimes \boldsymbol{\Phi}(i))\, E(\mathbf{u}_t^{\otimes 2} \otimes \mathbf{I}_n) + (\boldsymbol{\Sigma}^{\otimes 2}(i) \\
&\otimes \boldsymbol{\Phi}(i) \otimes \mathbf{c}(i))\, E(\mathbf{u}_t^{\otimes 2} \otimes \mathbf{I}_n) + (\boldsymbol{\Sigma}^{\otimes 2}(i) \otimes \mathbf{c}(i) \otimes \boldsymbol{\Phi}(i))\, E(\mathbf{u}_t^{\otimes 2} \otimes \mathbf{I}_n)]\, E(\mathbf{z}_{t-1} \,|\, s_t = i) \\
&+ \pi_i\ \mathbf{c}^{\otimes 4}(i)\ +\ \pi_i\, [\mathbf{c}^{\otimes 2}(i) \otimes \boldsymbol{\Sigma}^{\otimes 2}(i)\ +\ (\mathbf{c}(i) \otimes \boldsymbol{\Sigma}(i))^{\otimes 2}\ +\ \mathbf{c}(i) \otimes \boldsymbol{\Sigma}^{\otimes 2}(i) \otimes \mathbf{c}(i) \\
&+ \boldsymbol{\Sigma}(i) \otimes \mathbf{c}^{\otimes 2}(i) \otimes \boldsymbol{\Sigma}(i)\ +\ (\boldsymbol{\Sigma}(i) \otimes \mathbf{c}(i))^{\otimes 2}\ +\ \boldsymbol{\Sigma}^{\otimes 2}(i) \otimes \mathbf{c}^{\otimes 2}(i)]\, E(\mathbf{u}_t^{\otimes 2}) +\ \pi_i\, \boldsymbol{\Sigma}^{\otimes 4}(i)\, E(\mathbf{u}_t^{\otimes 4}).
\end{aligned}
$$

By Lemma 3.1, Equation (24) of Magnus and Neudecker (1986), and relations listed in the proof of Theorem 3.1, we have

$$
\begin{aligned}
E(\mathbf{I}_n^{\otimes 2} \otimes \mathbf{u}_t^{\otimes 2}) &= \mathbf{K}_{n^2,K^2}\ [\mathrm{vec}(\boldsymbol{\Omega}) \otimes \mathbf{I}_{n^2}] \\
E((\mathbf{I}_n \otimes \mathbf{u}_t)^{\otimes 2}) &= \mathbf{K}_{n^2,K^2}\ [\mathrm{vec}(\boldsymbol{\Omega}) \otimes \mathbf{I}_{n^2}] \\
E(\mathbf{I}_n \otimes \mathbf{u}_t^{\otimes 2} \otimes \mathbf{I}_n) &= \mathbf{K}_{n,nK^2}\ [\mathrm{vec}(\boldsymbol{\Omega}) \otimes \mathbf{I}_{n^2}]\, \mathbf{K}_{n,n} \\
E(\mathbf{u}_t \otimes \mathbf{I}_n^{\otimes 2} \otimes \mathbf{u}_t) &= \mathbf{K}_{n^2K,K}\ [\mathrm{vec}(\boldsymbol{\Omega}) \otimes \mathbf{I}_{n^2}] \\
E((\mathbf{u}_t \otimes \mathbf{I}_n)^{\otimes 2}) &= \mathbf{K}_{n,nK^2}\ [\mathrm{vec}(\boldsymbol{\Omega}) \otimes \mathbf{I}_{n^2}]\, \mathbf{K}_{n,n} \\
E(\mathbf{u}_t^{\otimes 2} \otimes \mathbf{I}_n^{\otimes 2}) &=\ \mathrm{vec}(\boldsymbol{\Omega}) \otimes \mathbf{I}_{n^2}.
\end{aligned}
$$

Using Lemma 3.1 and substituting these relations and those listed in the proof of Theorem 3.1 in the formula obtained above, we have

$$\pi_i\, E(\mathbf{z}_t^{\otimes 4}\,|\, s_t = i) = \sum_{j=1}^{M} p_{ji}\{\boldsymbol{\Phi}^{\otimes 4}(i)\}\ \pi_j\, E(\mathbf{z}_{t-1}^{\otimes 4}\,|\, s_{t-1} = j)$$

$$+ \sum_{j=1}^{M} p_{ji}\ \{\mathbf{c}(i) \otimes \boldsymbol{\Phi}^{\otimes 3}(i) + \boldsymbol{\Phi}(i) \otimes \mathbf{c}(i) \otimes \boldsymbol{\Phi}^{\otimes 2}(i) + \boldsymbol{\Phi}^{\otimes 2}(i) \otimes \mathbf{c}(i) \otimes \boldsymbol{\Phi}(i)$$

$$+ \boldsymbol{\Phi}^{\otimes 3}(i) \otimes \mathbf{c}(i)]\ \pi_j\, E(\mathbf{z}_{t-1}^{\otimes 3}\,|\, s_{t-1} = j)$$

$$+ \sum_{j=1}^{M} [\, p_{ji}\, \{\boldsymbol{\Phi}^{\otimes 2}(i) \otimes \mathbf{c}^{\otimes 2}(i) + (\boldsymbol{\Phi}(i) \otimes \mathbf{c}(i))^{\otimes 2} + \boldsymbol{\Phi}(i) \otimes \mathbf{c}^{\otimes 2}(i) \otimes \boldsymbol{\Phi}(i)$$

$$+ \mathbf{c}(i) \otimes \boldsymbol{\Phi}^{\otimes 2}(i) \otimes \mathbf{c}(i) + (\mathbf{c}(i) \otimes \boldsymbol{\Phi}(i))^{\otimes 2} + \mathbf{c}^{\otimes 2}(i) \otimes \boldsymbol{\Phi}^{\otimes 2}(i)\}$$

$$+\ p_{ji}\, \{(\boldsymbol{\Phi}^{\otimes 2}(i) \otimes \boldsymbol{\Sigma}^{\otimes 2}(i) + (\boldsymbol{\Phi}(i) \otimes \boldsymbol{\Sigma}(i))^{\otimes 2})\, \mathbf{K}_{n^2,K^2}$$

$$+ (\boldsymbol{\Sigma}(i) \otimes \boldsymbol{\Phi}^{\otimes 2}(i) \otimes \boldsymbol{\Sigma}(i))\, \mathbf{K}_{n^2K,K} + \boldsymbol{\Sigma}^{\otimes 2}(i) \otimes \boldsymbol{\Phi}^{\otimes 2}(i)\}\, (\mathrm{vec}(\boldsymbol{\Omega}) \otimes \mathbf{I}_{n^2})$$

$$+\ p_{ji}\, \{(\boldsymbol{\Phi}(i) \otimes \boldsymbol{\Sigma}^{\otimes 2}(i) \otimes \boldsymbol{\Phi}(i) + (\boldsymbol{\Sigma}(i) \otimes \boldsymbol{\Phi}(i))^{\otimes 2})\, \mathbf{K}_{n,nK^2}\}$$

$$\times\ (\mathrm{vec}(\boldsymbol{\Omega}) \otimes \mathbf{I}_{n^2})\, \mathbf{K}_{n,n}]\, \pi_j\, E(\mathbf{z}_{t-1}^{\otimes 2}\,|\, s_{t-1} = j)$$

$$+ \sum_{j=1}^{M} [\, p_{ji}\, \{\boldsymbol{\Phi}(i) \otimes \mathbf{c}^{\otimes 3}(i) + \mathbf{c}(i) \otimes \boldsymbol{\Phi}(i) \otimes \mathbf{c}^{\otimes 2}(i) + \mathbf{c}^{\otimes 2}(i) \otimes \boldsymbol{\Phi}(i) \otimes \mathbf{c}(i)$$

$$+ \mathbf{c}^{\otimes 3}(i) \otimes \boldsymbol{\Phi}(i)\} +\ p_{ji}\, \{(\boldsymbol{\Phi}(i) \otimes \mathbf{c}(i) \otimes \boldsymbol{\Sigma}^{\otimes 2}(i) + \boldsymbol{\Phi}(i) \otimes \boldsymbol{\Sigma}(i) \otimes \mathbf{c}(i) \otimes \boldsymbol{\Sigma}(i)$$

$$+ \boldsymbol{\Phi}(i) \otimes \boldsymbol{\Sigma}^{\otimes 2}(i) \otimes \mathbf{c}(i) + \mathbf{c}(i) \otimes \boldsymbol{\Phi}(i) \otimes \boldsymbol{\Sigma}^{\otimes 2}(i))\, \mathbf{K}_{n,K^2}$$

$$+ (\mathbf{c}(i) \otimes \boldsymbol{\Sigma}(i) \otimes \boldsymbol{\Phi}(i) \otimes \boldsymbol{\Sigma}(i) + \boldsymbol{\Sigma}(i) \otimes \boldsymbol{\Phi}(i) \otimes \mathbf{c}(i) \otimes \boldsymbol{\Sigma}(i) + \boldsymbol{\Sigma}(i) \otimes \boldsymbol{\Phi}(i) \otimes \boldsymbol{\Sigma}(i) \otimes \boldsymbol{c}(i)$$

$$+ \boldsymbol{\Sigma}(i) \otimes \mathbf{c}(i) \otimes \boldsymbol{\Phi}(i) \otimes \boldsymbol{\Sigma}(i))\, \mathbf{K}_{nK,K} + \mathbf{c}(i) \otimes \boldsymbol{\Sigma}^{\otimes 2}(i) \otimes \boldsymbol{\Phi}(i)\ + \boldsymbol{\Sigma}(i) \otimes \mathbf{c}(i) \otimes \boldsymbol{\Sigma}(i) \otimes \boldsymbol{\Phi}(i)$$

$$+ \boldsymbol{\Sigma}^{\otimes 2}(i) \otimes \boldsymbol{\Phi}(i) \otimes \mathbf{c}(i) + \boldsymbol{\Sigma}^{\otimes 2}(i) \otimes \mathbf{c}(i) \otimes \boldsymbol{\Phi}(i)\}\, (\mathrm{vec}(\boldsymbol{\Omega} \otimes \mathbf{I}_n)]\, \pi_j\, E(\mathbf{z}_{t-1}\,|\, s_{t-1} = j)$$

$$+\ \pi_i\ \ \{\mathbf{c}^{\otimes 4}(i)\} +\ \pi_i\, \{\mathbf{c}^{\otimes 2}(i) \otimes \boldsymbol{\Sigma}^{\otimes 2}(i) + (\mathbf{c}(i) \otimes \boldsymbol{\Sigma}(i))^{\otimes 2} + \mathbf{c}(i) \otimes \boldsymbol{\Sigma}^{\otimes 2}(i) \otimes \mathbf{c}(i)$$

$$+ \boldsymbol{\Sigma}(i) \otimes \mathbf{c}^{\otimes 2}(i) \otimes \boldsymbol{\Sigma}(i) + (\boldsymbol{\Sigma}(i) \otimes \mathbf{c}(i))^{\otimes 2} + \boldsymbol{\Sigma}^{\otimes 2}(i) \otimes \mathbf{c}^{\otimes 2}(i)\}\ \mathrm{vec}(\boldsymbol{\Omega})$$

$$+\ \pi_i\, \{\boldsymbol{\Sigma}^{\otimes 4}(i)\}\, \{\mathrm{vec}(\boldsymbol{\Omega} \otimes \boldsymbol{\Omega}) + \mathrm{vec}(\boldsymbol{\Omega}) \otimes \mathrm{vec}(\boldsymbol{\Omega}) + \mathrm{vec}[(\boldsymbol{\Omega} \otimes \boldsymbol{\Omega})\, \mathbf{K}_{K,K}]\}.$$

Reasoning as in the proof of Theorem 3.1, the last relation can be written as

$$\mathbf{H}_i = \sum_{j=1}^{M} \mathbf{P}(\mathbf{\Phi}^{\otimes 4})_{ij}\,\mathbf{H}_j \,+\, \sum_{j=1}^{M} \mathbf{X}_{ij}\,\mathbf{W}_j \,+\, \sum_{j=1}^{M} \mathbf{Y}_{ij}\,\mathbf{V}_j \,+\, \sum_{j=1}^{M} \mathbf{Z}_{ij}\,\mathbf{U}_j \,+\, \bar{\bar{\mathbf{C}}}_i$$
$$+\, \mathbf{R}_i\ \mathrm{vec}(\mathbf{\Omega}) \,+\, \bar{\bar{\mathbf{S}}}_i\,\{\mathrm{vec}(\mathbf{\Omega}\otimes\mathbf{\Omega}) + \mathrm{vec}(\mathbf{\Omega})\otimes\mathrm{vec}(\mathbf{\Omega}) + \mathrm{vec}[(\mathbf{\Omega}\otimes\mathbf{\Omega})\,\mathbf{K}_{K,K}]\}$$

where

$$\mathbf{Y}_{ij} = (\mathbf{Y}_1)_{ij} \,+\, (\mathbf{Y}_2)_{ij}\,(\mathrm{vec}(\mathbf{\Omega})\otimes\mathbf{I}_{n^2}) \,+\, (\mathbf{Y}_3)_{ij}\,(\mathrm{vec}(\mathbf{\Omega})\otimes\mathbf{I}_{n^2})\,\mathbf{K}_{n,n}$$
$$\mathbf{Z}_{ij} = (\mathbf{Z}_1)_{ij} \,+\, (\mathbf{Z}_2)_{ij}\,(\mathrm{vec}(\mathbf{\Omega})\otimes\mathbf{I}_n).$$

This proves Equation (4). Since $\rho(\mathbf{P}(\mathbf{\Phi}^{\otimes 4})) < 1$, the matrix $\mathbf{I}_{Mn^4} - \mathbf{P}(\mathbf{\Phi}^{\otimes 4})$ is invertible. So we can express $\mathbf{H}$ in closed form as in the statement of Theorem 3.2. Then we have $E(\mathbf{z}_t^{\otimes 4}) = \sum_{i=1}^{M} \pi_i\, E(\mathbf{z}_t^{\otimes 4}\,|\, s_t = i) = (\mathbf{e}^{'}\otimes \mathbf{I}_{n^4})\,\mathbf{H}$. The fourth moments of $(\mathbf{y}_t)$ are now easily deduced in matrix form.

4. Proof of Theorem 3.3

Set $r \geq 2$. To simplify computations, we assume that the process $(\mathbf{z}_t)$ is centered, that is, $\mathbf{c}_t = \mathbf{0}$. The result does not depend on the distribution of the residuals, hence they are assumed to be non-Gaussian i.i.d. with nontrivial finite moments at any dimension. In other words, we can choose residuals so that $\mathbf{A}(r)$ varies over the elements of a basis of $\mathbb{R}^{Mn^r}$. For every $k \geq 0$, let $\mathbf{A}(r,k)$ be the (Mn^r)-dimensional vector whose ith block is the n^r-dimensional vector $\pi_i\, E(\mathbf{z}_{t-k}\otimes\mathbf{z}_t^{\otimes(r-1)}\,|\, s_t = i)$ for $i = 1,\ldots,M$. For $k = 0$, we have $\mathbf{A}(r,k) = \mathbf{A}(r)$. First we prove that

$$\mathbf{A}(r,k) = \left[\mathbf{I}_n \otimes \mathbf{P}(\mathbf{\Phi}^{\otimes(r-1)})\right]\ \mathbf{A}(r,k-1).$$

If $r = 2$ and $\mathbf{c}_t = \mathbf{0}$, this formula is equivalent to $W(k) = P^* W(k-1)$ from FZ01, p.349, where $\mathbf{A}(2,k) = \operatorname{vec} W(k)$ and $\mathbf{P}(\mathbf{\Phi}) = P^*$, that is, $\operatorname{vec} W(k) = [\mathbf{I}_n \otimes P^*] \operatorname{vec} W(k-1)$. Using Result 1.1, we have

$$
\begin{aligned}
&\pi_i E(\mathbf{z}_{t-k} \otimes \mathbf{z}_t^{\otimes(r-1)} \mid s_t = i) = \pi_i \, E(\mathbf{z}_{t-k} \otimes (\mathbf{\Phi}(i)\,\mathbf{z}_{t-1} + \mathbf{\Sigma}(i)\,\mathbf{u}_t)^{\otimes(r-1)} \mid s_t = i) \\
&= \pi_i \sum_{j=1}^{M} \left[\mathbf{I}_n \otimes \mathbf{\Phi}^{\otimes(r-1)}(i)\right] \, E(\mathbf{z}_{t-k} \otimes \mathbf{z}_{t-1}^{\otimes(r-1)} \mid s_t = i, s_{t-1} = j)\, p_{ji}\, \pi_j \\
&= \sum_{j=1}^{M} \left[\mathbf{I}_n \otimes (p_{ji}\, \mathbf{\Phi}^{\otimes(r-1)}(i))\right] \, \pi_j \, E(\mathbf{z}_{t-k} \otimes \mathbf{z}_{t-1}^{\otimes(r-1)} \mid s_{t-1} = j)
\end{aligned}
$$

as $\mathbf{z}_{t-k}$ and $\mathbf{u}_t$ are uncorrelated for $k \geq 1$. This proves the above expression for $\mathbf{A}(r,k)$. By iteration, it gives

$$
\mathbf{A}(r,k) = \left[\mathbf{I}_n \otimes \{\mathbf{P}(\mathbf{\Phi}^{\otimes(r-1)})\}^k\right] \, \mathbf{A}(r)
$$

for every $k \geq 0$. Now we prove that

$$
\lim_{k \to +\infty} \mathbf{A}(r,k) \;=\; \mathbf{0}.
$$

From Result 1.1, we set

$$
\mathbf{z}_t = \sum_{\ell=0}^{+\infty} \mathbf{z}_{t,\ell} \qquad\qquad \mathbf{z}_{t,\ell} = \mathbf{\Phi}_t \mathbf{\Phi}_{t-1} \cdots \mathbf{\Phi}_{t-\ell+1} \mathbf{\Sigma}_{t-\ell} \mathbf{u}_{t-\ell}
$$

where $\mathbf{z}_{t,0} = \mathbf{\Sigma}_t \mathbf{u}_t$. Since $\mathbf{z}_{t,\ell}$ are uncorrelated and centered when $\mathbf{c}_t = \mathbf{0}$, we have

$$\pi_{i_1} E(\mathbf{z}_{t-k} \otimes \mathbf{z}_t^{\otimes(r-1)} \,|\, s_t = i_1) = \sum_{\ell_1=0}^{+\infty} \cdots \sum_{\ell_r=0}^{+\infty} \pi_{i_1} E(\mathbf{z}_{t-k,\ell_1} \otimes \mathbf{z}_{t,\ell_2} \otimes \cdots \otimes \mathbf{z}_{t,\ell_r} \,|\, s_t = i_1)$$
$$= \sum_{\ell_1=0}^{+\infty} \pi_{i_1} E(\mathbf{z}_{t-k,\ell_1} \otimes \mathbf{z}_{t,k+\ell_1}^{\otimes(r-1)} | s_t = i_1)$$

where $k + \ell_1 = \ell_2 = \cdots = \ell_r := \ell$. But we have

$$\pi_{i_1} E(\mathbf{z}_{t-k,\ell-k} \otimes \mathbf{z}_{t,\ell}^{\otimes(r-1)} \,|\, s_t = i_1) = \pi_{i_1} E(\mathbf{z}_{t-k,\ell-k} \otimes \mathbf{z}_{t,\ell}^{\otimes(r-1)} \,|\, s_t = i_1, s_{t-k} = i_1)\, p_{i_1 i_1}^{(k)}\, \pi_{i_1}$$
$$= \pi_{i_1}\, p_{i_1 i_1}^{(k)}\, \mathbf{\Phi}^{\otimes r}(i_1)\, E[(\mathbf{\Phi}_{t-k-1} \cdots \mathbf{\Phi}_{t-\ell+1} \mathbf{\Sigma}_{t-\ell} \mathbf{u}_{t-\ell})$$
$$\otimes (\mathbf{\Phi}_{t-1} \cdots \mathbf{\Phi}_{t-\ell+1} \mathbf{\Sigma}_{t-\ell} \mathbf{u}_{t-\ell})^{\otimes(r-1)} \,| s_t = i_1)$$
$$= \sum_{i_2=1}^{M} \pi_{i_1}\, p_{i_1 i_1}^{(k)}\, \mathbf{\Phi}^{\otimes r}(i_1)\, E[(\mathbf{\Phi}_{t-k-1} \cdots \mathbf{\Phi}_{t-\ell+1} \mathbf{\Sigma}_{t-\ell} \mathbf{u}_{t-\ell})$$
$$\otimes (\mathbf{\Phi}_{t-1} \cdots \mathbf{\Phi}_{t-\ell+1} \mathbf{\Sigma}_{t-\ell} \mathbf{u}_{t-\ell})^{\otimes(r-1)} | s_t = i_1, s_{t-1} = i_2) p_{i_2 i_1} \pi_{i_2}$$
$$= \sum_{i_2=1}^{M} p_{i_1 i_1}^{(k)}\, \mathbf{\Phi}^{\otimes r}(i_1)\, E[(\mathbf{\Phi}_{t-k-1} \cdots \mathbf{\Phi}_{t-\ell+1} \mathbf{\Sigma}_{t-\ell} \mathbf{u}_{t-\ell})$$
$$\otimes (\mathbf{\Phi}_{t-1} \cdots \mathbf{\Phi}_{t-\ell+1} \mathbf{\Sigma}_{t-\ell} \mathbf{u}_{t-\ell})^{\otimes(r-1)} | s_{t-1} = i_2, s_{t-k-1} = i_2)\, p_{i_2 i_2}^{(k)} p_{i_2 i_1} \pi_{i_2}\, \pi_{i_2}$$
$$= \sum_{i_2=1}^{M} p_{i_1 i_1}^{(k)}\, p_{i_2 i_2}^{(k)}\, \mathbf{\Phi}^{\otimes r}(i_1)\, \mathbf{\Phi}^{\otimes r}(i_2)\, E[(\mathbf{\Phi}_{t-k-2} \cdots \mathbf{\Phi}_{t-\ell+1} \mathbf{\Sigma}_{t-\ell} \mathbf{u}_{t-\ell})$$
$$\otimes (\mathbf{\Phi}_{t-2} \cdots \mathbf{\Phi}_{t-\ell+1} \mathbf{\Sigma}_{t-\ell} \mathbf{u}_{t-\ell})^{\otimes(r-1)} | s_{t-1} = i_2)\, p_{i_2 i_1} \pi_{i_2}$$

$$= \sum_{i_2=1}^{M} \sum_{i_3=1}^{M} p_{i_1 i_1}^{(k)} \, p_{i_2 i_2}^{(k)} \, \mathbf{\Phi}^{\otimes r}(i_1) \, \mathbf{\Phi}^{\otimes r}(i_2) \, E[(\mathbf{\Phi}_{t-k-2} \cdots \mathbf{\Phi}_{t-\ell+1} \mathbf{\Sigma}_{t-\ell} \mathbf{u}_{t-\ell})$$

$$\otimes (\mathbf{\Phi}_{t-2} \cdots \mathbf{\Phi}_{t-\ell+1} \mathbf{\Sigma}_{t-\ell} \mathbf{u}_{t-\ell})^{\otimes (r-1)} | s_{t-2} = i_3) \, p_{i_2 i_1} \, p_{i_3 i_2} \, \pi_{i_3}$$

$$\vdots$$

$$= \sum_{i_2=1}^{M} \cdots \sum_{i_{\ell_1+1}}^{M} p_{i_1 i_1}^{(k)} \cdots p_{i_{\ell_1} i_{\ell_1}}^{(k)} \, \mathbf{\Phi}^{\otimes r}(i_1) \cdots \mathbf{\Phi}^{\otimes r}(i_{\ell_1}) \, E[(\mathbf{\Sigma}_{t-\ell} \mathbf{u}_{t-\ell})$$

$$\otimes (\mathbf{\Phi}_{t-\ell_1} \cdots \mathbf{\Phi}_{t-\ell+1} \mathbf{\Sigma}_{t-\ell} \mathbf{u}_{t-\ell})^{\otimes (r-1)} | s_{t-\ell_1} = i_{\ell_1+1}) \, p_{i_2 i_1} \, p_{i_3 i_2} \cdots p_{i_{\ell_1+1} i_{\ell_1}} \, \pi_{i_{\ell_1+1}}$$

where the last expectation is a finite term. Here $p_{ij}^{(k)} = Pr(s_t = j | s_{t-k} = i)$ is the (i,j)th element of $\mathbf{P}^k$. The (i_1, i_{h+1}) block of the partitioned matrix $\{\mathbf{P}(\mathbf{\Phi}^{\otimes r})\}^h$ is exactly

$$\sum_{i_2=1}^{M} \cdots \sum_{i_{h+1}=1}^{M} p_{i_2 i_1} p_{i_3 i_2} \cdots p_{i_{h+1} i_h} \mathbf{\Phi}^{\otimes r}(i_1) \cdots \mathbf{\Phi}^{\otimes r}(i_h).$$

Now $\rho(\mathbf{P}(\mathbf{\Phi}^{\otimes r})) < 1$ if and only if $\lim_{h \to +\infty} \{\mathbf{P}(\mathbf{\Phi}^{\otimes r})\}^h = \mathbf{0}$. See, for example, Horn and Johnson (1985), Theorem 5.6.12, p.298. Thus we obtain

$$\lim_{h \to +\infty} \{\mathbf{P}(\mathbf{\Phi}^{\otimes r})\}_{i_1 i_{h+1}}^h = \mathbf{0}.$$

Using the above formula (with $h = \ell_1$) and $p_{i_j i_j}^{(k)} \to \pi_{i_j}$, for $k, \ell_1 \to +\infty$, we get

$$\lim_{k \to +\infty} \pi_{i_1} \, E(\mathbf{z}_{t-k} \otimes \mathbf{z}_t^{\otimes (r-1)} \, | \, s_t = i_1) = \mathbf{0}$$

for every $i_1 = 1, \ldots, M$. This proves that $\mathbf{A}(r,k)$ vanishes as k goes to infinity, hence

$$\lim_{k \to +\infty} [\mathbf{I}_n \otimes \{\mathbf{P}(\mathbf{\Phi}^{\otimes (r-1)})\}^k] \mathbf{A}(r) = \mathbf{0}.$$

From the above assumption on $\mathbf{A}(r)$, this implies

$$\lim_{k \to +\infty} \mathbf{I}_n \otimes \{\mathbf{P}(\mathbf{\Phi}^{\otimes(r-1)})\}^k = \mathbf{0}$$

that is, $\lim_{k \to +\infty} \{\mathbf{P}(\mathbf{\Phi}^{\otimes(r-1)})\}^k = \mathbf{0}$. Thus we have $\rho(\mathbf{P}(\mathbf{\Phi}^{\otimes(r-1)})) < 1$.

5. Proof of Corollary 3.1

By (5) we obtain the following linear system with unknown matrices $\mathbf{A}(1)$, $\ldots$, $\mathbf{A}(r)$:

$$\begin{aligned}
(\mathbf{I}_{Mn^r} - \mathbf{P}(\mathbf{\Phi}^{\otimes r}))\, \mathbf{A}(r) - \mathbf{B}_{r,1}\, \mathbf{A}(r-1) - \cdots - \mathbf{B}_{r,r-1}\, \mathbf{A}(1) &= \mathbf{B}_{r,r} \\
(\mathbf{I}_{Mn^{r-1}} - \mathbf{P}(\mathbf{\Phi}^{\otimes(r-1)})\, \mathbf{A}(r-1) - \cdots - \mathbf{B}_{r-1,r-2}\, \mathbf{A}(1) &= \mathbf{B}_{r-1,r-1} \\
&\vdots \\
(\mathbf{I}_{Mn} - \mathbf{P}(\mathbf{\Phi}))\, \mathbf{A}(1) &= \mathbf{B}_{1,1}.
\end{aligned}$$

The incomplete block matrix $\mathbf{\Lambda}$ associated to the system is $(M\delta) \times (M\delta)$, where $\delta = (n^{r+1} - n)/(n-1)$:

$$\mathbf{\Lambda} = \begin{pmatrix}
\mathbf{I}_{Mn^r} - \mathbf{P}(\mathbf{\Phi}^{\otimes r}) & -\mathbf{B}_{r,1} & \cdots & -\mathbf{B}_{r,r-1} \\
\mathbf{0} & \mathbf{I}_{Mn^{r-1}} - \mathbf{P}(\mathbf{\Phi}^{\otimes(r-1)}) & \cdots & -\mathbf{B}_{r-1,r-2} \\
\vdots & \vdots & & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_{Mn} - \mathbf{P}(\mathbf{\Phi})
\end{pmatrix}.$$

To determine explicitly the matrix expressions of every sth order moments of $(\mathbf{y}_t)$, the above system must satisfy the Cramer condition, that is,

$$\det(\mathbf{\Lambda}) = \prod_{s=1}^{r} \det(\mathbf{I}_{Mn^s} - \mathbf{P}(\mathbf{\Phi}^{\otimes s})) \neq 0$$

hence $\det(\mathbf{I}_{Mn^s} - \mathbf{P}(\mathbf{\Phi}^{\otimes s})) \neq 0$ for $s = 1, \ldots, r$. Thus $\mathbf{I}_{Mn^s} - \mathbf{P}(\mathbf{\Phi}^{\otimes s})$ must be invertible, for $s = 1, \ldots, r$. But this is ensured by the hypothesis $\rho(\mathbf{P}(\mathbf{\Phi}^{\otimes r})) < 1$ and Theorem 3.3.

6. Proof of Corollary 3.2

In this case $p_{ij} = Pr(s_t = j | s_{t-1} = i)$ does not depend on i, that is, $p_{ij} = \pi_j$. By the independence of the matrices $\mathbf{\Phi}_t^{\otimes r}$, we obtain

$$||E(\mathbf{z}_{t,\ell}^{\otimes r})|| = ||\{E(\mathbf{\Phi}_t^{\otimes r})\}^{\ell}\, E(\boldsymbol{\omega}_t^{\otimes r})|| \leq ||\{E(\mathbf{\Phi}_t^{\otimes r})\}^{\ell}||||E(\boldsymbol{\omega}_t^{\otimes r})|| < +\infty$$

where $|| \cdot ||$ is the L^2-norm. Reasoning as in the proof of Theorem 2 from FZ01, we can conclude that the sufficient condition in Theorem 3.4 can be replaced by $\rho(E(\mathbf{\Phi}_t^{\otimes r})) < 1$. See also Corollary 1 of Francq and Zakoïan (2005) for the case of MS GARCH models. But we have

$$E(\mathbf{\Phi}_t^{\otimes r}) = E[E(\mathbf{\Phi}_t^{\otimes r}|s_t)] = E[\sum_{i=1}^{M} \pi_i\, E(\mathbf{\Phi}_t^{\otimes r}|s_t = i)] = \sum_{i=1}^{M} \pi_i\, \mathbf{\Phi}^{\otimes r}(i).$$

This completes the proof.

7. Example 3.1

Set $\rho_r = \rho(\mathbf{P}(\mathbf{\Phi}^{\otimes r}))$. Then $\rho_r < 1$ is a sufficient condition for finite moments up to order r. This example shows that, in certain cases, it is not necessary. Let us consider Example 4 from FZ01, p.352, i.e., the univariate MS(2) AR(2) model defined by $y_t = \eta_t$ if $s_t = 1$ and $y_t = a y_{t-2} + \eta_t$ if $s_t = 2$. Here $\eta_t \sim IID(0, 1)$ and a is a nonzero real constant. As shown in FZ01, the process can be written as

$$y_t = \eta_t + \sum_{k=1}^{+\infty} a^k \, \eta_{t-2k} \, I_{s_t=2,\dots,s_{t-2k+2}=2}$$

where I_A is the indicator function of A. This implies that the necessary and sufficient condition for higher-order stationarity is simply $|a| < 1$. Then we have

$$\mathbf{\Phi}(1) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \qquad \mathbf{\Phi}(2) = \begin{pmatrix} 0 & a \\ 1 & 0 \end{pmatrix}.$$

Set $p_{11} = 0.1$ and $p_{22} = 0.2$. If $a = 1.08$, then $\rho_1 \sim 0.906$, $\rho_2 \sim 0.942$, $\rho_3 \sim 0.979$, and $\rho_4 \sim 1.017$. If $a = 1.2$, then $\rho_1 \sim 0.95$, $\rho_2 \sim 1.05$, $\rho_3 \sim 1.15$, and $\rho_4 \sim 1.26$. If $a = 1.039$, then $\rho_7 \sim 0.9967$ and $\rho_8 \sim 1.0159$. This shows that the converse of Theorem 3.3 is not true in general. One can see that $\rho_r = |a|^{r/2} \sqrt{p_{12}\, p_{21} + p_{22}^2}$, hence the parameter restrictions in terms of "parameter region" is defined by $|a|^{r/2} \sqrt{p_{12}\, p_{21} + p_{22}^2} < 1$, $p_{12} = 1 - p_{11}$, $p_{21} = 1 - p_{22}$, $0 < p_{11} < 1$, $0 < p_{22} < 1$, $|a| > 0$, and $r \geq 1$. This proves that condition $\rho_r < 1$ can be unnecessary. If the chain is i.i.d., then the sufficient condition becomes $|a|^{r/2} \sqrt{\pi_2} < 1$.

8. Proof of Theorem 4.2

i) Kollo and Srivastava (2004) proved that $\beta_1^{[Ma]}(\mathbf{y})$ can be expressed via the third order multivariate moments as follows

$$\beta_1^{[Ma]}(\mathbf{y}) = \text{trace}(E((\mathbf{y}_t^*)' \otimes \mathbf{y}_t^* \otimes (\mathbf{y}_t^*)')\, E(\mathbf{y}_t^* \otimes (\mathbf{y}_t^*)' \otimes \mathbf{y}_t^*)).$$

Then we have

$$\begin{aligned}\beta_1^{[Ma]}(\mathbf{y}) &= [\text{vec}(E(\mathbf{y}_t^* \otimes (\mathbf{y}_t^*)' \otimes \mathbf{y}_t^*))]' \text{ vec}(E(\mathbf{y}_t^* \otimes (\mathbf{y}_t^*)' \otimes \mathbf{y}_t^*)) \\ &= [(\mathbf{K}_{K,K} \otimes \mathbf{I}_K)\, E(\mathbf{y}_t^* \otimes \mathbf{y}_t^* \otimes \mathbf{y}_t^*)]' (\mathbf{K}_{K,K} \otimes \mathbf{I}_K)\, E(\mathbf{y}_t^* \otimes \mathbf{y}_t^* \otimes \mathbf{y}_t^*) \\ &= E((\mathbf{y}_t^*)' \otimes (\mathbf{y}_t^*)' \otimes (\mathbf{y}_t^*)')\, (\mathbf{K}_{K,K} \otimes \mathbf{I}_K)\, (\mathbf{K}_{K,K} \otimes \mathbf{I}_K)\, E(\mathbf{y}_t^* \otimes \mathbf{y}_t^* \otimes \mathbf{y}_t^*) \\ &= E((\mathbf{y}_t^*)' \otimes (\mathbf{y}_t^*)' \otimes (\mathbf{y}_t^*)')\, E(\mathbf{y}_t^* \otimes \mathbf{y}_t^* \otimes \mathbf{y}_t^*) = (\mathbf{s}^{[C]}(\mathbf{y}))'\, \mathbf{s}^{[C]}(\mathbf{y}).\end{aligned}$$

From Kollo (2008), p.2333, the following relation

$$\beta_2^{[Ma]}(\mathbf{y}) = \text{trace}(E(\mathbf{y}_t^*\, (\mathbf{y}_t^*)' \otimes \mathbf{y}_t^*\, (\mathbf{y}_t^*)'))$$

holds. Then we get

$$\begin{aligned}\beta_2^{[Ma]}(\mathbf{y}) &= E(\text{trace}(\mathbf{I}_{K^2}\, (\mathbf{y}_t^*\, (\mathbf{y}_t^*)' \otimes \mathbf{y}_t^*\, (\mathbf{y}_t^*)'))) \\ &= E([\text{vec}(\mathbf{I}_{K^2})]' \text{ vec}(\mathbf{y}_t^*\, (\mathbf{y}_t^*)' \otimes \mathbf{y}_t^*\, (\mathbf{y}_t^*)')) \\ &= [\text{vec}(\mathbf{I}_{K^2})]'\, E(\mathbf{y}_t^* \otimes \mathbf{y}_t^* \otimes \mathbf{y}_t^* \otimes \mathbf{y}_t^*) = [\text{vec}(\mathbf{I}_{K^2})]'\, \mathbf{k}^{[C]}(\mathbf{y}).\end{aligned}$$

ii) By Móri *et al.* (1993) we have $\mathbf{s}^{[MRS]}(\mathbf{y}) = E(||\mathbf{y}_t^*||^2\, \mathbf{y}_t^*)$ hence

$$\begin{aligned}\mathbf{s}^{[MRS]}(\mathbf{y}) &= E((\mathbf{y}_t^*)'\, \mathbf{y}_t^*\, \mathbf{y}_t^*) = \text{vec}\, E((\mathbf{y}_t^*)'\, \mathbf{y}_t^*\, \mathbf{y}_t^*) \\ &= E(\text{vec}((\mathbf{y}_t^*)'\, \mathbf{y}_t^*\, \mathbf{y}_t^*)) = E(\text{vec}(\mathbf{y}_t^*\, (\mathbf{y}_t^*)'\, \mathbf{y}_t^*)) \\ &= E(\text{vec}(\mathbf{y}_t^*\, (\mathbf{y}_t^*)'\, \mathbf{I}_K\, \mathbf{y}_t^*)) = E([(\mathbf{y}_t^*)' \otimes (\mathbf{y}_t^*\, (\mathbf{y}_t^*)')] \text{ vec}\, \mathbf{I}_K) \\ &= E((\mathbf{y}_t^*)' \otimes (\mathbf{y}_t^*\, (\mathbf{y}_t^*)')) \text{ vec}\, \mathbf{I}_K = \text{vec}\{E((\mathbf{y}_t^*)' \otimes (\mathbf{y}_t^*\, (\mathbf{y}_t^*)')) \text{ vec}\, \mathbf{I}_K\} \\ &= ([\text{vec}(\mathbf{I}_K)]' \otimes \mathbf{I}_K) \text{ vec}\, E((\mathbf{y}_t^*)' \otimes (\mathbf{y}_t^*\, (\mathbf{y}_t^*)')) \\ &= ([\text{vec}(\mathbf{I}_K)]' \otimes \mathbf{I}_K)\, E(\text{vec}(\mathbf{y}_t^*)' \otimes \text{vec}(\mathbf{y}_t^*\, (\mathbf{y}_t^*)')) \\ &= ([\text{vec}(\mathbf{I}_K)]' \otimes \mathbf{I}_K)\, E(\mathbf{y}_t^* \otimes \mathbf{y}_t^* \otimes \mathbf{y}_t^*) = ([\text{vec}(\mathbf{I}_K)]' \otimes \mathbf{I}_K)\, \mathbf{s}^{[C]}(\mathbf{y}).\end{aligned}$$

Furthermore, we have

$$\begin{aligned}
\operatorname{vec}\mathbf{K}^{[MRS]}(\mathbf{y}) &= \operatorname{vec} E(\mathbf{y}_t^*\,(\mathbf{y}_t^*)'\,\mathbf{y}_t^*(\mathbf{y}_t^*)') - (K+2)\operatorname{vec}\mathbf{I}_K \\
&= E(\operatorname{vec}(\mathbf{y}_t^*\,(\mathbf{y}_t^*)'\,\mathbf{y}_t^*(\mathbf{y}_t^*)')) - (K+2)\operatorname{vec}\mathbf{I}_K \\
&= E((\mathbf{y}_t^*\,(\mathbf{y}_t^*)') \otimes (\mathbf{y}_t^*\,(\mathbf{y}_t^*)'))\ \operatorname{vec}\mathbf{I}_K\ - (K+2)\operatorname{vec}\mathbf{I}_K \\
&= ([\operatorname{vec}\mathbf{I}_K]' \otimes \mathbf{I}_{K^2})\, E(\operatorname{vec}((\mathbf{y}_t^*\,(\mathbf{y}_t^*)') \otimes (\mathbf{y}_t^*\,(\mathbf{y}_t^*)'))) - (K+2)\operatorname{vec}\mathbf{I}_K \\
&= ([\operatorname{vec}\mathbf{I}_K]' \otimes \mathbf{I}_{K^2})\, E(\mathbf{y}_t^* \otimes \mathbf{y}_t^* \otimes \mathbf{y}_t^* \otimes \mathbf{y}_t^*) - (K+2)\operatorname{vec}\mathbf{I}_K \\
&= ([\operatorname{vec}\mathbf{I}_K]' \otimes \mathbf{I}_{K^2})\, \mathbf{k}^{[C]}(\mathbf{y}) - (K+2)\operatorname{vec}\mathbf{I}_K.
\end{aligned}$$

iii) From Kollo (2008), p.2332, we obtain

$$\mathbf{b}^{[Ko]}(\mathbf{y}) = \mathbf{1}_{K\times K} \star E(\mathbf{y}_t^* \otimes (\mathbf{y}_t^*)' \otimes \mathbf{y}_t^*) = E[\sum_{i,j} (y_{ti}^*\, y_{tj}^*)\, \mathbf{y}_t^*]$$

hence

$$\begin{aligned}
&\mathbf{b}^{[Ko]}(\mathbf{y}) = E[(\mathbf{y}_t^*)'\,\mathbf{1}_{K\times K}\,\mathbf{y}_t^*\,\mathbf{y}_t^*] = E(\operatorname{vec}(\mathbf{y}_t^*\,(\mathbf{y}_t^*)'\,\mathbf{1}_{K\times K}\,\mathbf{y}_t^*)) \\
&= E([(\mathbf{y}_t^*)' \otimes (\mathbf{y}_t^*\,(\mathbf{y}_t^*)')]\ \operatorname{vec}\mathbf{1}_{K\times K}) = \operatorname{vec}\{\mathbf{I}_K\, E([(\mathbf{y}_t^*)' \otimes (\mathbf{y}_t^*\,(\mathbf{y}_t^*)')]\,\mathbf{1}_{K^2\times 1}\} \\
&= (\mathbf{1}_{1\times K^2} \otimes \mathbf{I}_K)\ \operatorname{vec} E((\mathbf{y}_t^*)' \otimes (\mathbf{y}_t^*\,(\mathbf{y}_t^*)')) = (\mathbf{1}_{1\times K^2} \otimes \mathbf{I}_K)\,\mathbf{s}^{[C]}(\mathbf{y}).
\end{aligned}$$

Finally, we have

$$\begin{aligned}
&\operatorname{vec}\mathbf{B}^{[Ko]}(\mathbf{y}) = \operatorname{vec} E((\mathbf{y}_t^*)'\,\mathbf{1}_{K\times K}\,\mathbf{y}_t^*\,\mathbf{y}_t^*\,(\mathbf{y}_t^*)') = E((\mathbf{y}_t^*)'\,\mathbf{1}_{K\times K}\,\mathbf{y}_t^*\,\mathbf{y}_t^* \otimes \mathbf{y}_t^*) \\
&= E(\operatorname{vec}((\mathbf{y}_t^*)'\,\mathbf{1}_{K\times K}\,\mathbf{y}_t^*)\,\mathbf{y}_t^* \otimes \mathbf{y}_t^*) = E((\mathbf{y}_t^*)' \otimes (\mathbf{y}_t^*)'\ \operatorname{vec}(\mathbf{1}_{K\times K})\,\mathbf{y}_t^* \otimes \mathbf{y}_t^*) \\
&= E((\mathbf{y}_t^* \otimes \mathbf{y}_t^*)\,((\mathbf{y}_t^*)' \otimes (\mathbf{y}_t^*)')\ \operatorname{vec}(\mathbf{1}_{K\times K})) = E((\mathbf{y}_t^* \otimes \mathbf{y}_t^*)\,((\mathbf{y}_t^*)' \otimes (\mathbf{y}_t^*)'))\ \operatorname{vec}(\mathbf{1}_{K\times K}) \\
&= E((\mathbf{y}_t^*\,(\mathbf{y}_t^*)') \otimes (\mathbf{y}_t^*\,(\mathbf{y}_t^*)'))\ \operatorname{vec}(\mathbf{1}_{K\times K}) = \operatorname{vec}[\mathbf{I}_{K^2}\, E((\mathbf{y}_t^*\,(\mathbf{y}_t^*)') \otimes (\mathbf{y}_t^*\,(\mathbf{y}_t^*)'))\,\mathbf{1}_{K^2\times 1}] \\
&= (\mathbf{1}_{1\times K^2} \otimes \mathbf{I}_{K^2})\ \operatorname{vec} E((\mathbf{y}_t^*\,(\mathbf{y}_t^*)') \otimes (\mathbf{y}_t^*\,(\mathbf{y}_t^*)')) = (\mathbf{1}_{1\times K^2} \otimes \mathbf{I}_{K^2})\,\mathbf{k}^{[C]}(\mathbf{y}).
\end{aligned}$$

9. Proof of Corollary 4.1

By Kollo (2008), p.2334, we get

$$\operatorname{trace}\mathbf{B}^{[Ko]}(\mathbf{y}_t) = \sum_{i,j,k} E(y^*_{ti}\, y^*_{tj}\, (y^*_{tk})^2) = E((\sum_{i,j=1}^{K} y^*_{ti}\, y^*_{tj})\, (\sum_{k=1}^{K} (y^*_{tk})^2))$$

$$= E(((\mathbf{y}^*_t)'\, \mathbf{1}_{K\times K}\, \mathbf{y}^*_t)\, ||\mathbf{y}^*_t||^2) = E(((\mathbf{y}^*_t)'\, \mathbf{1}_{K\times K}\, \mathbf{y}^*_t)\, (\mathbf{y}^*_t)'\, \mathbf{y}^*_t)$$

$$= E((\mathbf{y}^*_t)'\, \mathbf{y}^*_t\, (\mathbf{y}^*_t)'\, \mathbf{1}_{K\times K}\, \mathbf{y}^*_t) = E[(\mathbf{1}_{K\times K}\, \mathbf{y}^*_t)' \otimes (\mathbf{y}^*_t)'\ \operatorname{vec}(\mathbf{y}^*_t\, (\mathbf{y}^*_t)')]$$

$$= E[((\mathbf{y}^*_t)'\, \mathbf{1}_{K\times K}) \otimes (\mathbf{y}^*_t)'\, \mathbf{y}^*_t \otimes \mathbf{y}^*_t] = E[((\mathbf{y}^*_t)' \otimes (\mathbf{y}^*_t)')\, (\mathbf{1}_{K\times K} \otimes \mathbf{I}_K)\, \mathbf{y}^*_t \otimes \mathbf{y}^*_t]$$

$$= E[(\mathbf{y}^*_t)' \otimes (\mathbf{y}^*_t)' \otimes (\mathbf{y}^*_t)' \otimes (\mathbf{y}^*_t)']\ \operatorname{vec}(\mathbf{1}_{K\times K} \otimes \mathbf{I}_K) = (\mathbf{k}^{[C]}(\mathbf{y}))'\ \operatorname{vec}(\mathbf{1}_{K\times K} \otimes \mathbf{I}_K)$$

and

$$||\mathbf{B}^{[Ko]}(\mathbf{y})||^2 = [\operatorname{vec}\mathbf{B}^{[Ko]}(\mathbf{y})]'\ [\operatorname{vec}\mathbf{B}^{[Ko]}(\mathbf{y})]$$

$$= [(\mathbf{1}_{1\times K^2} \otimes \mathbf{I}_{K^2})\mathbf{k}^{[C]}(\mathbf{y})]'\ [(\mathbf{1}_{1\times K^2} \otimes \mathbf{I}_{K^2})\mathbf{k}^{[C]}(\mathbf{y})]$$

$$= \mathbf{k}^{[C]}(\mathbf{y})'\ (\mathbf{1}_{K^2\times K^2} \otimes \mathbf{I}_{K^2})\, \mathbf{k}^{[C]}(\mathbf{y}) = \operatorname{trace}[(\mathbf{1}_{K^2\times K^2} \otimes \mathbf{I}_{K^2})\, \mathbf{k}^{[C]}(\mathbf{y})\, \mathbf{k}^{[C]}(\mathbf{y})']$$

$$= [\operatorname{vec}(\mathbf{1}_{K^2\times K^2} \otimes \mathbf{I}_{K^2})]'\ \operatorname{vec}(\mathbf{k}^{[C]}(\mathbf{y})\, \mathbf{k}^{[C]}(\mathbf{y})')$$

$$= [\operatorname{vec}(\mathbf{1}_{K^2\times K^2} \otimes \mathbf{I}_{K^2})]'\, \mathbf{k}^{[C]}(\mathbf{y}) \otimes \mathbf{k}^{[C]}(\mathbf{y})$$

where $\mathbf{1}_{1\times K^2} \otimes \mathbf{I}_{K^2} = (\mathbf{I}_{K^2}\ \cdots\ \mathbf{I}_{K^2}) = \mathbf{I} \in \mathbb{R}^{K^2\times K^4}$.

10. Example 4.1

Let us consider the bivariate ($K = 2$) simulated model $\mathbf{y}_t = c(s_t) + \sigma(s_t)\, \mathbf{u}_t$, $\mathbf{u}_t \sim NID(\mathbf{0}, \mathbf{I}_2)$ with $M = 3$. We set the following parameters:

$$c(1) = [1 \quad 2]' \qquad c(2) = [0.2 \quad 0.5]' \qquad c(3) = [-1 \quad 0]'$$

$$\sigma(1) = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \qquad \sigma(2) = \begin{bmatrix} 3 & 2 \\ 2 & 4 \end{bmatrix} \qquad \sigma(3) = \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix}.$$

The transition probability matrix and the ergodic probability vector are:

$$\mathcal{P} = \begin{bmatrix} 0.60 & 0.02 & 0.02 \\ 0.20 & 0.90 & 0.08 \\ 0.20 & 0.08 & 0.90 \end{bmatrix} \qquad \boldsymbol{\pi} = \begin{bmatrix} 0.0476 \\ 0.4762 \\ 0.4762 \end{bmatrix}.$$

Firstly, we compute the four measures of skewness as presented above:

$$\mathbf{s}^{[C]} = [0.2613 \ -0.1152 \ -0.1152 \ -0.1751 \ -0.1152 \ -0.1751 \ -0.1751 \ 1.0900]'$$
$$\beta_1^{[Ma]} = 1.3881$$
$$\mathbf{s}^{[MRS]} = [0.0862 \quad 0.9747]'$$
$$\mathbf{b}^{[Ko]} = [-0.1443 \quad 0.6246]'.$$

Then we report the measures of multivariate kurtosis:

$$\mathbf{k}^{[C]} = [10.2201 \ -14.5945 \ -14.5945 \ 22.5217 \ -14.5945 \ 22.5217 \ 22.5217 \ -36.9711 \\ -14.5945 \ 22.5217 \ 22.5217 \ -36.9711 \ 22.5217 \ -36.9711 \ -36.9711 \ 65.4624]'$$
$$\beta_2^{[Ma]} = 120.7259$$
$$\text{vec}\ \mathbf{K}^{[MRS]} = [28.7418 \quad -51.5656 \quad -51.5656 \quad 83.9841]'$$
$$\text{vec}\ \mathbf{B}^{[Ko]} = [3.5528 \quad -6.5222 \quad -6.5222 \quad 14.0420]'.$$

All those measures correctly detect the leptokurtic and almost symmetric characteristics of the process. We see that the departure from normality of this process is substantial as we compare these numerical vectors with those of a normal population $\mathbf{w}$ of dimension $K = 2$: $\beta_2^{[Ma]}(\mathbf{w}) = 8$, $\mathbf{K}^{[MRS]}(\mathbf{w}) = \mathbf{0}$, $\mathbf{B}^{[Ko]}(\mathbf{w}) = \text{vec}[4\ 2\ 2\ 4]'$ and $\mathbf{k}^{[C]}(\mathbf{w}) = [3\ 0\ 0\ 1\ 0\ 1\ 1\ 0\ 0\ 1\ 1\ 0\ 1\ 0\ 0\ 3]'$.

Note that the various definitions of skewness for $\mathbf{w}$ are all null.

11. Example 4.2

Let us consider the bivariate MS(2) VARMA(1, 1) model defined in the online Supplementary Material, Section 2. The measures of skewness for this switching model clearly indicate a symmetric distribution:

$$\mathbf{s}^{[C]} = [0\ 0\ 0\ 0\ 0\ 0\ 0\ 0]' \qquad \beta_1^{[Ma]} = 0 \quad \mathbf{s}^{[MRS]} = [0 \qquad 0]' \qquad \mathbf{b}^{[Ko]} = [0 \qquad 0]'.$$

However, the measures of kurtosis vary as follows:

$$\begin{aligned}\mathbf{k}^{[C]} = [&4.1402 \quad 0.0180 \quad -0.0857 \quad 1.2531 \quad -0.1212 \quad 0.8223 \quad 1.2467 \quad -0.0526 \\ &-0.0175 \quad 1.2612 \quad 0.8223 \quad -0.0762 \quad 1.2549 \quad -0.3059 \quad -0.2823 \quad 9.5463]'\end{aligned}$$

$$\beta_2^{[Ma]} = 15.3312 \qquad \text{vec}\ \mathbf{K}^{[MRS]} = [1.3951 \quad -0.2879 \quad -0.3679 \quad 6.7994]'$$

$$\text{vec}\ \mathbf{B}^{[Ko]} = [5.2564 \quad 1.7957 \quad 1.7011 \quad 10.6706]'.$$

Now we formally test the departure from normality. By Example 4.1 above, we have $||\mathbf{k}^{[C]}||^2 - ||\mathbf{k}^{[C]}(\mathbf{w})||^2 = 116 - 24 = 92$, where $\mathbf{w}$ is normally distributed. This leads to a rejection of the null of normality. The rejection is also confirmed by Mardia's measure as $\beta_2^{[Ma]} - \beta_2^{[Ma]}(\mathbf{w}) = 15.33 - 8 = 7.33$ is greater than the critical values of a normal distribution at any given significance level (see Theorem 4.2). Finally, by Corollary 4.1, we have $||\mathbf{B}^{[Ko]}||^2 - ||\mathbf{B}^{[Ko]}(\mathbf{w})||^2 = 147 - 40 = 107$. This again rejects normality. Although all measures of skewness agree on a symmetric feature of the process, the testing procedures on kurtosis' measures clearly indicate presence of excess kurtosis with respect to the normal behaviour.

References

Costa, O.L.V., Fragoso, M.D., Marques, R.P., Discrete-Time Markov Jump Linear Systems. in: Probability and Its Applications (Gani, J., Heyde, C.C., Jagers, P., and Kurtz, T.G., eds.), Springer–Verlag, London, 2005.

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

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

Horn, R.A., Johnson, C.R., Matrix Analysis, Cambridge Univ. Press, Cambridge, 1985.

Kollo, T., Multivariate skewness and kurtosis measures with an application in ICA, Journal of Multivariate Analysis **99** (2008), 2328–2338.

Kollo, T., and Srivastava, M.S., Estimation and testing of parameters in multivariate Laplace distribution, Comm. Statist. **33** (2004), 2363–2687.

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

Mardia, K.V., Measures of multivariate skewness and kurtosis with applications, Biometrika **57** (1970), 519–530.

Móri, T.F., Rohatgi, V.K., and Székely, G.J., On multivariate skewness and kurtosis, Theory Probab. Appl. **38** (1993), 547–551.