Accepted Manuscript

Skewness and kurtosis of multivariate Markov-switching processes

Gabriele Fiorentini, Christophe Planas, Alessandro Rossi

PII: S0167-9473(15)00147-4
DOI: http://dx.doi.org/10.1016/j.csda.2015.06.009
Reference: COMSTA 6105

To appear in: Computational Statistics and Data Analysis

Received date: 19 December 2014
Revised date: 25 June 2015
Accepted date: 25 June 2015

Please cite this article as: Fiorentini, G., Planas, C., Rossi, A., Skewness and kurtosis of multivariate Markov-switching processes. Computational Statistics and Data Analysis (2015), http://dx.doi.org/10.1016/j.csda.2015.06.009

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Skewness and kurtosis of multivariate Markov-switching processes*

Gabriele Fiorentini†, Christophe Planas‡ and Alessandro Rossi‡

First version: November 2014
This version: June 2015

Abstract

Exact formulae are provided for the calculation of multivariate skewness and kurtosis of Markov-switching Vector Auto-Regressive (MS VAR) processes as well as for the general class of MS state space (MS SS) models. The use of the higher-order moments in non-linear modelling is illustrated with two examples. A Matlab code that implements the results is available from the authors.

Keywords: Markov-switching; VAR models; Higher-order moments; State space models.

*E-Mails: fiorentini@disia.unifi.it, christophe.planas@jrc.ec.europa.eu, alessandro.rossi@jrc.ec.europa.eu
†University of Florence, School of Economics, and Rimini Centre for Economic Analysis
‡European Commission, Joint Research Centre
1 Introduction

Markov-switching models are now widespread in applied macroeconomics and finance. By extending linear specifications with a discrete latent process that controls parameter switches, MS models have gained the ability to fit time series subject to non-linearities. In macroeconomics, MS models have been introduced by Hamilton (1989) with the aim of capturing the asymmetry of the business cycle. Kim and Nelson (1999a) and Mc Connell and Perez-Quiros (2000) have extended the Hamilton model to account for the reduction in business cycle fluctuations known as the Great Moderation. Phillips (1991) has applied the Hamilton model to a multi-country case. Ang and Beckaert (2002a) have underlined the usefulness of a multivariate dimension when analyzing switches in the dynamics of the US, UK, and German short-term interest rates. Favero and Monacelli (2005) and Sims and Zha (2006) have resorted to the MS VAR framework to detect shifts in the US monetary and fiscal policy. Given the empirical evidence about the existence of policy regimes, the last generation of dynamic stochastic general equilibrium models includes Markov-switching policy reaction functions (see Davig and Leeper, 2007, and Davig, Leeper, and Chung, 2004). In this context MS VAR models arise as fundamental solution of the forward-looking structural equations (Farmer, Waggoner, and Zha, 2009, 2011). MS models have also been intensively used in empirical finance to reproduce the fat tails, leverage effects, volatility clustering, and time-varying correlations that characterize many financial return series. Also in this context switching regimes have been inserted into equilibrium models: Cecchetti, Lam, and Mark (1990, 1993), for instance, have added regimes to the conventional asset pricing model through switching processes for dividends and consumption. General discussions and additional references can be found in Krolzig (1997), Kim and Nelson (1999b), Fruhwirth-Schnatter (2006), Ang and Timmermann (2011), and Guidolin (2012).

The statistical properties of MS VAR models have been analyzed, among others, by Yang (2000), Francq and Zakoian (2001, 2002), and Cavicchioli (2013, 2014). These studies focus on stationarity issues, on the first two unconditional moments, and on the determination of the number of regimes. Timmermann (2000) derives the first four moments for univariate MS models. In spite of their relevance, the higher-order moments are still unknown in the general multivariate case. We give closed-form formulae for multivariate MS VAR processes as well as for the general class of MS state space models (see Kim 1993, 1994). In an independent research paper, Cavicchioli (2015) also derives closed-form expressions for the moments of MS VARMA models up to any order and proposes alternative measures of skewness and kurtosis.
The general MS VAR and MS state space models are presented in Section 2. We focus on models where the discrete latent variable takes a finite number of states with time-invariant transition probabilities. In Section 3, we derive formulae for the higher-order moments for both MS VAR and MS SS models. The use of the higher-order moments is illustrated in Section 4 with two examples. Section 5 concludes.

2 Model and assumptions

Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space on which a vector process \(\{\varepsilon_t\}\) and a \(K\)-state Markov chain \(\{S_t\}\) are defined at discrete time \(t\). The first-order MS VAR process for the \(n_x\)-dimensional vector \(\{x_t\}\) is generated by the stochastic difference equation:

\[
x_t = \alpha_{S_t} + \Phi_{S_t} x_{t-1} + \Lambda_{S_t} \varepsilon_t
\]  

(2.1)

Specifications involving more lags can easily be cast into the formulation above through the VAR(1) companion form. The following assumptions are supposed to hold:

(i) The \(n_x\)-dimensional shocks \(\{\varepsilon_t\}\) verify \(\varepsilon_t \sim iid N(0, I_{n_x})\), where \(I_{n_x}\) denotes the \(n_x \times n_x\) identity matrix.

(ii) The Markov chain \(\{S_t\}\) is homogeneous, irreducible, aperiodic, and non-null persistent. These conditions ensure stationarity of \(\{S_t\}\) (see e.g. Grimmett and Stirzaker, 1992).

(iii) The processes \(\{\varepsilon_t\}\) and \(\{S_t\}\) are independent.

(iv) The \(n_x \times 1\) vector \(\alpha_{S_t}\), and the \(n_x \times n_x\) matrices \(\Phi_{S_t}\) and \(\Lambda_{S_t}\) take as many values as the realization of \(S_t\), i.e. \(\alpha_{S_t} \in \{\alpha_1, \ldots, \alpha_K\}\), \(\Phi_{S_t} \in \{\Phi_1, \ldots, \Phi_K\}\), and \(\Lambda_{S_t} \in \{\Lambda_1, \ldots, \Lambda_K\}\).

The variable \(x_t\) may be unobserved. In this case it is typically related to a vector of \(n_y\) observations \(y_t\) through the measurement equation:

\[
y_t = a_{S_t} + H_{S_t} x_t + \gamma_{S_t} u_t
\]  

(2.2)

Equations (2.1)-(2.2) make up a MS state space model. The following additional assumptions are made:

(v) The \(n_u\)-dimensional process \(\{u_t\}\) is such that \(u_t \sim iid N(0, I_{n_u})\).
(vi) The process \( \{u_t\} \) is independent of \( \{\varepsilon_t\} \) and \( \{S_t\} \).

(vii) The \( n_y \times 1 \) vector \( a_{S_t} \), the \( n_y \times n_x \) matrix \( H_{S_t} \), and the \( n_y \times n_u \) matrix \( \gamma_{S_t} \) take \( K \) different values depending on the realization of the discrete latent variable \( S_t \).

Throughout the paper, we denote \( p_{jk} \) the conditional probability \( p_{jk} = P(S_t = k | S_{t-1} = j) \) for \( j, k = 1, \cdots, K \), \( \pi_k \) the marginal probability of state \( k \) i.e. \( \pi_k = P(S_t = k) \), and \( \pi \) the \( K \times K \) matrix with \( (\pi_1, \cdots, \pi_K) \) on the main diagonal and zeros elsewhere. We call \( J_{n,k} \) the \( n \times nK \) matrix \( J_{n,k} = [0_n \times n_k (k-1), I_n, 0_n \times n (K-k)] \), \( k = 1, \cdots, K \); all together the \( K \) matrices \( J_{n,k} \) sum to \( J_n = \sum_{k=1}^{K} J_{n,k} \). For any two \( n \times 1 \) vectors \( A \) and \( B \) we denote \( R_n \) the \( n^2 \times n^2 \) commutation matrix such that \( A \otimes B = R_n (B \otimes A) \). This commutation matrix can be built as \( R_n = \sum_{i=1}^{n} \sum_{j=1}^{n} (e_i e_j') \otimes (e_j e_i') \), \( e_j \) being the \( n \times 1 \) canonical vector \( e_j = [0_{j-1}, 1, 0_{n-j}]' \) (see Magnus and Neudecker, 1999).

For any integer \( m \), we also define \( P_m(\Phi) \) the \( Kn^m \times Kn^m \) matrix such that:

\[
P_m(\Phi) = \begin{pmatrix}
\Phi_1 \otimes \cdots \otimes \Phi_1 P_{11} & \Phi_1 \otimes \cdots \otimes \Phi_1 P_{21} & \cdots & \Phi_1 \otimes \cdots \otimes \Phi_1 P_{1K} \\
\Phi_2 \otimes \cdots \otimes \Phi_2 P_{12} & \Phi_2 \otimes \cdots \otimes \Phi_2 P_{22} & \cdots & \Phi_2 \otimes \cdots \otimes \Phi_2 P_{2K} \\
\vdots & \vdots & \ddots & \vdots \\
\Phi_K \otimes \cdots \otimes \Phi_K P_{1K} & \Phi_K \otimes \cdots \otimes \Phi_K P_{2K} & \cdots & \Phi_K \otimes \cdots \otimes \Phi_K P_{KK}
\end{pmatrix}
\]

(2.3)

Finally we denote \( \rho(M) \) the spectral radius of matrix \( M \).

3 Multivariate measures of skewness and kurtosis

In macroeconomics and finance, non-linearities are typically analyzed through pairwise measures of skewness and kurtosis such as:

\[
\frac{Cov(w_{it}^{k}, w_{jt}^{\ell})}{V(w_{it})^{k/2}V(w_{jt})^{\ell/2}}
\]

(3.1)

where \( w_{it} \) and \( w_{jt} \) are two scalar elements of a \( n \)-dimensional random vector \( w_t \) with finite moments, \( V \) and \( Cov \) stand for variance and covariance respectively, and \( k, \ell \) are strictly positive integers whose sum \( k + \ell = 3, 4 \) gives the moment order. The case \( i = j \) yields the univariate higher order moments and \( i \neq j \) the mixed-moments. All moments given by (3.1) can be
collected into the $n^3 \times 1$ and $n^4 \times 1$ vectors:

$$
Sk(w_t; \Sigma_w) = E[\Sigma_w^{-1/2}(w_t - E(w_t)) \otimes \Sigma_w^{-1/2}(w_t - E(w_t)) \otimes \Sigma_w^{-1/2}(w_t - E(w_t))] \\
Ku(w_t; \Sigma_w) = E[\Sigma_w^{-1/2}(w_t - E(w_t)) \otimes \Sigma_w^{-1/2}(w_t - E(w_t)) \otimes \Sigma_w^{-1/2}(w_t - E(w_t)) \otimes \Sigma_w^{-1/2}(w_t - E(w_t))]
$$

(3.2)

where $E$ is the expectation and $\Sigma_w$ is variance-covariance matrix of $w_t$ with zeros outside the main diagonal, i.e. $\Sigma_w = diag[V(w_t)]$.

Mardia (1970) proposed alternative definitions of multivariate skewness and kurtosis known as $\beta_{1,n}$ and $\beta_{2,n}$ which aggregate univariate and mix-moments. As shown in Kollo and Srivastava (2005) and Kollo (2008), Mardia’s statistics can be easily retrieved from (3.2) since:

$$
\beta_{1,n} = tr\{E[w_t^* \otimes w_t^* \otimes w_t^*]E[w_t^* \otimes w_t^* \otimes w_t^*]\} = Sk(w_t; \Lambda_w)Sk(w_t; \Lambda_w) \\
\beta_{2,n} = tr\{E[w_t^* w_t^* \otimes w_t^* w_t^*]\} = Vec(I_{n^2})Ku(w_t; \Lambda_w)
$$

where $tr$ represents trace, $w_t^*$ is such that $w_t^* = \Lambda_w^{-1/2}(w_t - E(w_t))$, and $\Lambda_w$ is any symmetric square root of $V(w_t)$. Other measures of skewness and kurtosis can be found in the literature, for instance in Mori, Rohatgi, and Szekely (1993); they can be similarly calculated using formula (3.2). Without loss of generality we focus on the measures of skewness and kurtosis given in (3.2).

### 3.1 Markov-switching vector autoregressive models

Given the MS VAR process $\{x_t\}$ in (2.1) with assumptions (i)-(iv), let us define $z_t$ the standardized variable $z_t = \Sigma_x^{-1/2}(x_t - E(x_t))$ where $\Sigma_x = diag[V(x_t)]$. By construction $Sk(x_t; \Sigma_x) = E(z_t \otimes z_t \otimes z_t)$ and $Ku(x_t; \Sigma_x) = E(z_t \otimes z_t \otimes z_t)$. It is easily checked that $\{z_t\}$ follows the MS VAR process:

$$
z_t = c_{S_t} + \Psi_{S_t} z_{t-1} + \Omega_{S_t} \epsilon_t \\
c_{S_t} = \Sigma_x^{-1/2}(\alpha_{S_t} - E(x_t)) + \Phi_{S_t} E(x_t) \\
\Psi_{S_t} = \Sigma_x^{-1/2} \Phi_{S_t} \Sigma_x^{1/2} \\
\Omega_{S_t} = \Sigma_x^{-1/2} \Lambda_{S_t}
$$

(3.3)
Theorem 1 below uses the auxiliary process \( \{z_t\} \) for deriving the skewness and kurtosis for the general MS VAR process (2.1).

**Theorem 1**  Suppose \( \{x_t\} \) follows the process (2.1) and that 

\[
\rho(P_m(\Phi)) < 1 \text{ for } m \leq 4.
\]

Then:

(I) the skewness of the vector \( x_t \) is given by

\[
Sk(x_t; \Sigma_x) = J_{n_x^3}[I_{Kn_x^3} - P_3(\Psi)]^{-1}D
\]

where \( D \) is the \( Kn_x^3 \times 1 \) vector

\[
d_k = c_k \otimes c_k \otimes c_k + A_3\{(\Omega_k \otimes \Omega_k \otimes c_k)\text{Vec}(I_{n_x})
\]

\[
+ \left[c_k \otimes c_k + (\Omega_k \otimes \Omega_k)\text{Vec}(I_{n_x})\right] \otimes \left\{\Psi_k \sum_{j=1}^K p_{jk} \frac{\pi_j}{\pi_k} \Sigma_x^{-1/2} [E(x_t|S_t = j) - E(x_t)]\right\}
\]

\[
+ \left(\Psi_k \otimes \Psi_k \otimes c_k\right) \sum_{j=1}^K p_{jk} \frac{\pi_j}{\pi_k} (\Sigma_x^{-1/2} \otimes \Sigma_x^{-1/2})
\]

\[
\times \left[E(x_t \otimes x_t|S_t = j) - E(x_t|S_t = j) \otimes E(x_t) - E(x_t) \otimes E(x_t|S_t = j) + E(x_t) \otimes E(x_t)\right]\}
\]

\(c_k, \Psi_k\) and \(\Omega_k\) are defined in (3.3), the \( Kn_x^3 \times Kn_x^3 \) matrix \( P_3(\Psi) \) is like in (2.3), \( \Sigma_x = \text{diag}[V(x_t)] \), and the \( n_x^3 \times n_x^3 \) matrix \( A_3 \) is given below

\[
A_3 = I_{n_x^3} + (I_{n_x} \otimes R_{n_x}) + (R_{n_x} \otimes I_{n_x})(I_{n_x} \otimes R_{n_x}),
\]

(II) the kurtosis of the vector \( x_t \) is given by

\[
Ku(x_t; \Sigma_x) = J_{n_x^4}[I_{Kn_x^4} - P_4(\Psi)]^{-1}M
\]

where \( P_4(\Psi) \) is like in (2.3), \( M \) is the \( Kn_x^4 \times 1 \) vector \( M = (\pi_1 m_1, \pi_2 m_2, \ldots, \pi_K m_K)' \) whose \( n_x^4 \times 1 \) elements \( m_k \) are equal to:
\[ m_k = c_k \otimes c_k \otimes c_k \otimes c_k + A_4(c_k \otimes c_k \otimes \Omega_k \otimes \Omega_k)Vec(I_{n_x}) + (\Omega_k \otimes \Omega_k \otimes \Omega_k \otimes \Omega_k)B \]

\[ + \tilde{A}_4\left\{ [c_k \otimes c_k \otimes c_k + A_3(\Omega_k \otimes \Omega_k \otimes c_k)Vec(I_{n_x})] \otimes \left( \Psi_k \sum_{j=1}^{K} p_{jk} \frac{\pi_j}{\pi_k} \Sigma_x^{-1/2}[E(x_t|S_t = j) - E(x_t)] \right) \right\} \]

\[ + A_4[\Psi_k \otimes \Psi_k \otimes c_k \otimes c_k + (\Psi_k \otimes \Psi_k \otimes \Omega_k \otimes \Omega_k)Vec(I_{n_x})] \sum_{j=1}^{K} p_{jk} \frac{\pi_j}{\pi_k} (\Sigma_x^{-1/2} \otimes \Sigma_x^{-1/2}) \]

\[ \times [E(x_t \otimes x_t|S_t = j) - E(x_t|S_t = j) \otimes E(x_t) - E(x_t) \otimes E(x_t)] \]

\[ + \tilde{A}_4(\Psi_k \otimes \Psi_k \otimes \Psi_k \otimes c_k) \sum_{j=1}^{K} p_{jk} J_{n_x, j}^2 [I_{Kn_x^2} - P_3(\Psi)]^{-1} D, \]

where the matrix \( D \) is detailed in (I) above, the \( n_x^4 \times n_x^4 \) matrix \( A_4 \) and the \( n_x^4 \)-dimensional vector \( B \) verify:

\[ A_4 = I_{n_x^4} + (I_{n_x} \otimes R_{n_x} \otimes I_{n_x})(I_{n_x} \otimes R_{n_x} \otimes I_{n_x}) \]

\[ + (I_{n_x^2} \otimes R_{n_x})(I_{n_x} \otimes R_{n_x} \otimes I_{n_x}) \]

\[ + (I_{n_x} \otimes R_{n_x} \otimes I_{n_x})(I_{n_x^2} \otimes R_{n_x})(R_{n_x} \otimes I_{n_x}) \]

\[ + (I_{n_x} \otimes R_{n_x} \otimes I_{n_x})(I_{n_x^2} \otimes R_{n_x})(I_{n_x} \otimes R_{n_x} \otimes I_{n_x}) \]  

(3.7)

\[ B = Vec(I_{n_x^2} + R_{n_x}) + Vec(I_{n_x}) \otimes Vec(I_{n_x}) \]  

(3.8)

and the \( n_x^4 \times n_x^4 \) matrix \( \tilde{A}_4 \) is such as

\[ \tilde{A}_4 = I_{n_x^4} + (I_{n_x^2} \otimes R_{n_x}) + (I_{n_x} \otimes R_{n_x} \otimes I_{n_x})(I_{n_x^2} \otimes R_{n_x}) \]

\[ + (R_{n_x} \otimes I_{n_x^2})(I_{n_x} \otimes R_{n_x} \otimes I_{n_x})(I_{n_x^2} \otimes R_{n_x}) \]  

(3.9)

**Proof:** See Supplementary material.

The conditional moments \( E(x_t|S_t = j) \) and \( E(x_t \otimes x_t|S_t = j), j = 1, 2, \cdots, K, \) are given in Lemma 1 in Supplementary material. It is easily checked that the relationship \( \Psi_{S_t} = \Sigma_x^{-1/2} \Phi_{S_t} \Sigma_x^{-1/2} \) and the assumptions \( \rho(P_3(\Phi)) < 1 \) and \( \rho(P_4(\Phi)) < 1 \) imply invertibility of the matrices \( I_{Kn_x^2} - P_3(\Psi) \) and \( I_{Kn_x^4} - P_4(\Psi) \) in equations (3.4) and (3.6).
In the absence of the autoregressive lag, i.e. when $\Phi S_t = 0$, model (2.1) is gaussian conditionally to the concurrent state $S_t$ so the distribution of $x_t$ is a finite mixture of normal densities (see for example Fiorentini, Planas, and Rossi, 2014). In empirical finance the efficient market hypothesis provides a compelling argument for excluding autoregressive terms, so the finite mixture model has often been applied to the analysis of returns, for instance by Ang and Beckaert (2002b) and Taamouti (2012). Theorem 1 simplifies as follows:

$$ Sk(x_t; \Sigma_x) = \sum_{k=1}^{K} \pi_k \{ c_k \otimes c_k \otimes c_k + A_3(\Omega_k \otimes \Omega_k \otimes c_k)V ec(I_{n_x}) \} $$

$$ Ku(x_t; \Sigma_x) = \sum_{k=1}^{K} \pi_k \{ c_k \otimes c_k \otimes c_k \otimes c_k + A_4(c_k \otimes c_k \otimes \Omega_k \otimes \Omega_k)V ec(I_{n_x}) $$

$$ + (\Omega_k \otimes \Omega_k \otimes \Omega_k \otimes \Omega_k)B \} \tag{3.10} $$

We turn to the moments of MS SS process.

### 3.2 Markov-switching state space models

The first two unconditional moments of vector $y_t$ in (2.1)-(2.2) are easily derived from the state conditional moments $E(x_t|S_t)$ and $E(x_t \otimes x_t|S_t)$ since:

$$ E(y_t) = \sum_{k=1}^{K} \pi_k [a_k + H_k E(x_t|S_t = k)] $$

$$ V ec[V(y_t)] = \sum_{k=1}^{K} \pi_k [a_k \otimes a_k + H_k \otimes H_k E(x_t \otimes x_t|S_t = k) + (\gamma_k \otimes \gamma_k)V ec(I_{n_x}) $$

$$ + (a_k \otimes H_k + H_k \otimes a_k)E(x_t|S_t = k)] - E(y_t) \otimes E(y_t) \tag{3.11} $$

Like for the MS VAR case, we define $y_t^*$ the standardized variable $y_t^* = \Sigma_y^{-1/2}(y_t - E(y_t))$ where $\Sigma_y = diag[V(y_t)]$. Again, this standardization simplifies algebra since $Sk(y_t; \Sigma_y) = E(y_t^* \otimes y_t^* \otimes y_t^*)$ and $Ku(y_t; \Sigma_y) = E(y_t^* \otimes y_t^* \otimes y_t^* \otimes y_t^*)$. It is easily checked that $\{y_t^*\}$ follows
the process:

\[ y_t^* = a_{St}^* + H_{St}^* z_t + \gamma_{St}^* u_t \]

\[ a_{St}^* = \Sigma_y^{-1/2}(a_{St} - E(y_t) + H_{St}E(x_t)) \]

\[ H_{St}^* = \Sigma_y^{-1/2}H_{St}\Sigma_x^{1/2} \]

\[ \gamma_{St}^* = \Sigma_y^{-1/2}\gamma_{St} \]  \hspace{1cm} (3.12)

where \( z_t \) and \( \Sigma_x \) are defined in Section 3.1. Theorem 2 below provides the skewness and kurtosis of MS SS processes.

**Theorem 2** Suppose \( \{y_t\} \) evolves as in (2.1)-(2.2) and that \( \rho(P_m(\Phi)) < 1 \) for \( m \leq 4 \). Then:

(I) the skewness of the vector \( y_t \) is given by

\[ Sk(y_t; \Sigma_y) = \sum_{k=1}^{K} \pi_k \left\{ a_k^* \otimes a_k^* \otimes a_k^* + A_3^* \left[ (\gamma_k^* \otimes \gamma_k^* \otimes a_k^*) Vec(I_{n_y}) + (a_k^* \otimes a_k^* + \gamma_k^* \otimes \gamma_k^* Vec(I_{n_y})) \otimes (H_k^* \Sigma_x^{-1/2}[E(x_t|S_t = k) - E(x_t)]) \right] \right. \]

\[ + \left. (H_k^* \otimes H_k^* \otimes a_k^*) (\Sigma_x^{-1/2} \otimes \Sigma_x^{-1/2}) \times (E(x_t \otimes x_t|S_t = k) - E(x_t|S_t = k) \otimes E(x_t) - E(x_t) \otimes E(x_t|S_t = k) + E(x_t) \otimes E(x_t)) \right] \}

\[ + (H_k^* \otimes H_k^* \otimes H_k^*) J_{n^2_y,3} [I_{Kn^2_y} - P_3(\Psi)]^{-1} D \]  \hspace{1cm} (3.13)

where \( a_k^* \), \( H_k^* \), and \( \gamma_k^* \) are shown in (3.12), the \( n_y \times n_y \) matrix \( A_3^* \) is like in (3.5) with dimension \( n_y \) instead of \( n_x \), \( \Sigma_x = \text{diag}[V(x_t)] \), and \( D \) is detailed in Theorem 1.
(II) the kurtosis of the vector \( y_t \) is given by

\[
Ku(y_t; \Sigma_y) = \sum_{k=1}^{K} \pi_k \{ a_k^* \otimes a_k^* \otimes a_k^* \otimes a_k^* + A_k^*(a_k^* \otimes a_k^* \otimes \gamma_k^* \otimes \gamma_k^*) \operatorname{Vec}(I_{n_u}) + (\gamma_k^* \otimes \gamma_k^* \otimes \gamma_k^* \otimes \gamma_k^*) B^* \}
\]

+ \( \tilde{A}_4^*(a_k^* \otimes a_k^* \otimes a_k^* \otimes H_k^*) \Sigma_x^{-1/2} \left[ E(x_t|S_t = k) - E(x_t) \right] \)

+ \( A_4^*(a_k^* \otimes H_k^* \otimes \gamma_k^* \otimes \gamma_k^*) \left[ \Sigma_x^{-1/2} \left[ E(x_t|S_t = k) - E(x_t) \right] \otimes \operatorname{Vec}(I_{n_u}) \right] \)

+ \( A_4^*(a_k^* \otimes a_k^* \otimes H_k^* \otimes H_k^*) \)

\( \times \left( \Sigma_x^{-1/2} \otimes \Sigma_x^{-1/2} \right) \left[ E(x_t \otimes x_t|S_t = k) - E(x_t \otimes x_t|S_t = k) \otimes E(x_t) \right] \)

+ \( E(x_t) \otimes E(x_t) + A_4^*(H_k^* \otimes H_k^* \otimes \gamma_k^* \otimes \gamma_k^*) \)

\( \times \left[ \operatorname{Vec}(I_{n_u}) \otimes (\Sigma_x^{-1/2} \otimes \Sigma_x^{-1/2}) \left[ E(x_t \otimes x_t|S_t = k) - E(x_t \otimes x_t|S_t = k) \otimes E(x_t) \right] \right. \)

- \( E(x_t) \otimes E(x_t|S_t = k) + E(x_t) \otimes E(x_t) \}

+ \( \tilde{A}_4^*(H_k^* \otimes H_k^* \otimes H_k^* \otimes a_k^*) J_{n^3_x,k} [I_{Kn^3_u} - P_3(\Psi)]^{-1} D \)

+ \( (H_k^* \otimes H_k^* \otimes H_k^* \otimes H_k^*) J_{n^3_x,k} [I_{Kn^3_u} - P_4(\Psi)]^{-1} M \)  \( (3.14) \)

where \( M \) is detailed in Theorem 1, the \( n_y^4 \times n_y^4 \) matrices \( A_4^* \) and \( \tilde{A}_4^* \) are like in (3.7) and (3.9) with dimension \( n_y \) instead of \( n_x \), and the \( n_u^4 \times 1 \) vector \( B^* \) is like in (3.8) with dimension \( n_u \) instead of \( n_x \).

The proof is omitted as it follows closely that of Theorem 1 when \( \Phi_{S_i} = 0 \), the measurement equation (2.2) not involving autoregressive lags. It makes use of the higher-order moments of the state variable \( x_t \) which are known. Two examples below illustrate the use of the higher-order moments in multivariate MS models.

4 Examples

UK asset returns

Guidolin and Timmermann (GT, 2005) fit a MS VAR model to the UK stock and bond monthly excess returns for the period 1976-2 to 2000-12. They consider three regimes that impact the intercept, the autoregressive matrix, and the shocks variance-covariance matrix. The regimes are interpreted as bear, normal, and bull market periods. Table 1 shows the model-based skewness and kurtosis of the UK stock and bond excess returns as implied by the parameter
values given in GT’s Table 4. The co-skewness statistics reported in Table 1 relates the level of the first variable to the square of the second one, whereas the co-kurtosis relates the level of the first variable to the cube of the second variable. In order to gauge the model fit, the empirical counterparts are also displayed together with the 95% confidence intervals computed using the block bootstrap proposed by Politis and Romano (1994).

Table 1: Higher-order moments of UK stock and bond excess returns

|        | Skewness | Kurtosis | Co-skewness | Co-kurtosis |
|--------|----------|----------|-------------|-------------|
| **Stocks** |          |          |             |             |
| Empirical | -1.17    | 8.68     | 0.11        | 2.05        |
|          | (-2.12,-0.06) | (3.21, 13.37) | (-0.20, 0.40) | (0.90, 3.10) |
| Model-based | -0.53    | 6.92     | -0.02       | 2.65        |
| **Bonds**  |          |          |             |             |
| Empirical  | 0.55     | 6.08     | 0.15        | -0.02       |
|           | (-0.30, 1.34) | (3.09, 8.88) | (-0.28, 0.55) | (-2.71, 2.30) |
| Model-based | -0.18    | 5.08     | -0.05       | 3.15        |

Notes: the model-based moments have been calculated using the parameter estimates given in Table 4 of GT (2005); the co-skewness relates the level of the variable of interest to the square of the other one; the co-kurtosis relates the level of the variable of interest to the cube of the other variable; 95% confidence intervals are reported between parenthesis.

The empirical excess kurtosis are significantly greater than zero, justifying the use of a non-linear model for describing UK stock and bond returns. The two univariate model-based kurtosis confirms that the model correctly weights extreme returns on the two assets. The sample skewness of stocks is significantly negative while no asymmetry is detected in the distribution of bond returns. The model adequately captures these features. The empirical co-skewness of stock and bond returns are almost null, suggesting that the level of each variable is not impacted by the volatility of the other one. The model catches this feature also remarkably well. With a value equal to 2.05, the empirical co-kurtosis of stock returns suggests that extreme values of bond returns have some impact on average stock returns. Conversely, the empirical co-kurtosis of bond returns is almost null: extreme values of stock returns have no impact on bond returns on average. With a value equal to 3.15, the model-based co-kurtosis of bonds
falls outside the empirical confidence interval, so the model leads to the unsupported conclusion that bonds cannot diversify the risk inherent to a portfolio of stocks.

**US business cycle**

Chauvet (1998) and Kim and Nelson (1998) consider a MS dynamic factor model to extract a composite index of the US business cycle out of the growth rates of four US macroeconomic series, namely industrial production, non-farm payroll employment, personal income less transfer payments, and real manufacturing and trade sales. The dynamic factor model is specified as:

\[
M_0 : \quad y_{it} = \lambda_i f_t + \nu_{it}
\]

\[
v_{it} = \phi_{i1} v_{it-1} + \phi_{i2} v_{it-2} + \sigma_i \epsilon_{it}
\]

\[
f_t = \mu S_{1t} + a_t
\]

where \(a_t\) and \(\epsilon_{it}\), \(i = 1, \ldots, 4\), are standard gaussian white noises. The mean of the common factor \(f_t\) switches between two values \(\mu_1, \mu_2\), according to the phase of the business cycle which is indexed by the discrete latent variable \(S_{1t} \in \{1,2\}\). Camacho, Perez-Quiros, and Poncela (CPP, 2012), estimate model \(M_0\) on US monthly observations from January 1967 to November 2010.

Table 2 shows both empirical and model-based moments of the growth rate of Industrial Production and Employment. For the two variables the empirical skewness is negative as well as the co-skewness. The model adequately reproduces these features. The two series exhibit an excess kurtosis which is sizeable and significant. Model \(M_0\) however implies almost zero excess kurtosis. Table 2 also shows the co-kurtosis statistics which relates the square of the two variables. Its empirical value is equal to 3.57 with confidence interval (2.34, 5.21). Since the theoretical value under normality equals 1.37, this reveals the presence of excess co-movements in volatility between Employment and Industrial Production in the US. Model \(M_0\) however does not foresee this feature as it implies a co-kurtosis of 1.62, outside of the confidence interval. To catch this non-linearity, we allow for heteroskedasticity in the common shock \(a_t\) as in:

\[
M_1 : \quad f_t = \mu S_{1t} + \sigma_{S_{2t}} a_t
\]

The variance of \(\sigma_{S_{2t}} a_t\) now switches between two regimes according to the two-state Markov-variable \(S_{2t}\) which is independent of \(S_{1t}\).
Table 2: Higher-order moments of US industrial production and employment

|           | IP          | EM          | EM - IP     |
|-----------|-------------|-------------|-------------|
|           | Skewness    | Kurtosis    | Skewness    | Kurtosis    | Co-skewness | Co-kurtosis |
| Empirical | -0.93       | 6.99        | -0.43       | 5.05        | -0.69       | 3.57        |
|           | (-1.48, -0.06) | (4.13, 9.17) | (-1.01, 0.23) | (3.65, 7.14) | (-1.10, -0.14) | (2.34, 5.21) |
| Model-based | M₀ | -0.31       | 3.24        | -0.10       | 3.05        | -0.22       | 1.62        |
|           | M₁ | -0.08       | 4.15        | -0.06       | 3.79        | -0.07       | 2.53        |

Notes: IP refers to the US Industrial Production Index and EM to the US non-farm employment; co-skewness refers to the third-moment that involves employment and the square of Industrial Production; the co-kurtosis statistics relates the square of the two variables; M₀: model with switching growth; M₁: model with switches in growth and in volatility; 95% confidence intervals are reported between parenthesis.

We estimate model M₁ by approximated maximum likelihood (Kim, 1994). The higher-order moments under M₁ are displayed in the last row of Table 2. Model M₁ yields third and fourth moments that lie inside the empirical 95% confidence intervals. Hence modelling co-movements in volatility improves the characterization of US Employment and Industrial Production compared to the original CPP’s specification.

5 Conclusion

We extend the early work by Timmermann (2000) on univariate MS models by deriving closed-form formulae for the multivariate skewness and kurtosis in both MS VAR and MS state space models. Besides enriching the model interpretation by summarizing non-linear features, these formulae provide a useful tool for diagnostic checking via moment-matching. A Matlab code that implements the results in the paper is available from the authors.
Acknowledgments

We are grateful to Dante Amengual, Gaby Perez-Quiros and Enrique Sentana as well as audiences at CEMFI Madrid, JRC Ispra, RCEF 2014 Rimini, and ESEM 2014 Toulouse for their comments and suggestions. Special thanks are due to Massimo Guidolin for very helpful discussions. We also thank the Associate Editor and two anonymous referees for useful comments that have greatly help us improving the paper. Of course, the usual caveat applies. Fiorentini acknowledges funding from MIUR PRIN MISURA - Multivariate models for risk assessment. The views expressed in this paper are those of the authors and should not be attributed to the European Commission.

References

ANG A. AND BEKAERT G. (2002a), “Regime switches in interest rates”, Journal of Business & Economic Statistics, 20, 163-182.

ANG A. AND BEKAERT G. (2002b), “International asset allocation with regime shifts”, Review of Financial Studies, 15, 4, 1137-1187.

ANG A. AND TIMMERMANN A. (2011), “Regime changes and financial markets”, National Bureau of Economic Research, Working Paper 17182.

CAVICCHIOLI M. (2013), “Spectral density of Markov-switching VARMA models”, Economic Letters, 121, 2, 218-220.

CAVICCHIOLI M. (2014), “Determining the number of regimes in Markov switching VAR and VMA models”, Journal of Time Series Analysis, 35, 2, 173-186.

CAVICCHIOLI M. (2015), “Higher order moments of Markov switching VARMA models”, mimeo.

CECCHETTI. S., LAM P. S., AND MARK N. (1990), “Mean reversion in equilibrium asset prices”, American Economic Review, 80, 398-41.

CECCHETTI. S., LAM P. S., AND MARK N. (1993), “The equity premium and the risk-free rate”, Journal of Monetary Economics, 31, 21-45.

CHAUVE T. (1998), “An econometric characterization of business cycle dynamics with factor structure and regime switching”, International Economic Review, 39, 4, 969-996.
Davig T., and Leeper E.M. (2007), “Generalizing the Taylor principle”, *American Economic Review*, 97, 3, 607-635.

Davig T., Leeper E. M., and Chung H. (2004), “Monetary and fiscal policy switching”, National Bureau of Economic Research, Working Paper 10362.

Farmer R. E., Waggoner D. F., and Zha T. (2009), “Understanding Markov-switching rational expectations models”, *Journal of Economic Theory*, 144, 1849-1867.

Farmer R. E., Waggoner D. F., and Zha T. (2011), “Minimal state variable solutions to Markov-switching rational expectations models”, *Journal of Economic Dynamics & Control*, 35, 2150-2166.

Favero C., and Monacelli T. (2005), “Fiscal policy rules and regime (in)stability: evidence from the U.S.”, Innocenzo Gasparini Institute for Economic Research, Working Paper 282.

Fiorentini G., Planas C., and Rossi A. (2014), “Marginal distribution of Markov-switching VAR processes”, mimeo.

Francq C., and Zakoian J. M. (2001), “Stationarity of multivariate Markov-switching ARMA models”, *Journal of Econometrics*, 102, 339-364.

Francq C., and Zakoian J. M. (2002), “Comments on the paper by Minxian Yang: some properties of vector autoregressive processes with Markov-switching coefficients”, *Econometric Theory*, 18, 815-818.

Fruhwirth-Schnatter S. (2006), *Finite Mixture and Markov Switching Models*, New York, Springer.

Grimmett G. R., and Stirzaker D. R. (1992), *Probability and Random Processes*, Oxford, Clarendon Press, 2nd edition.

Guidolin M. (2012), “Markov switching models in empirical finance”, Innocenzo Gasparini Institute for Economic Research, Working Paper no. 415.

Guidolin M., and Timmermann A. (2005), “Economic implications of bull and bear regimes in UK stock and bond returns”, *Economic Journal*, 115, 111-143.

Hamilton J. D. (1989), “A new approach to the economic analysis of nonstationary time series and the business cycle”, *Econometrica*, 57, 2, 357-384.

Kim C. J. (1993), “Unobserved component time series models with Markov-switching heteroskedasticity: changes in regimes and the link between inflation rates and inflation uncer-
Kim C. J. (1994), “Dynamic linear models with Markov-switching”, Journal of Econometrics, 60, 1-22.

Kim C. J., and Nelson C. R. (1998), “Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching”, Review of Economics and Statistics, 188-201.

Kim C. J., and Nelson C. R. (1999a), “Has the US economy become more stable: a Bayesian approach based on a Markov-switching model of the business cycle”, Review of Economics and Statistics, 81, 608-616.

Kim C. J., and Nelson C. R. (1999b), State-space Models with Regime Switching: Classical and Gibbs Sampling Approaches with Applications, Massachusetts Institute of Technology Press, Cambridge.

Kollo T. (2008), “Multivariate skewness and kurtosis measures with an application to ICA”, Journal of Multivariate Analysis, 99, 2328-2338.

Kollo T., and Srivastava M. S. (2005), “Estimation and testing of parameters in multivariate Laplace distribution”, Communications in Statistics - Theory and Methods, 33, 2363-2387.

Krolzig H. M. (1997), Markov-Switching Vector Autoregressions: Modelling, Statistical Inference, and Application to Business Cycle Analysis, Lecture Notes in Economics and Mathematical Systems, New York, Springer.

Magnus J., and Neudecker H. (1999), Matrix Differential Calculus With Applications in Statistics and Econometrics, John Wiley & Sons Ltd, Chichester.

Mardia K. V. (1970), “Measures of multivariate skewness and kurtosis with applications”, Biometrika, 57, 3, 519-530.

Mc Connell M. M., and Perez-Quiros G. (2000), “Output fluctuations in the United States: what has changed since the early 1980’s”, American Economic Review, 90, 5, 1464-1476.

Mori T. F., Rohatgi V. K., and Szekely G. J. (1993), “On multivariate skewness and kurtosis”, Theory of Probability and Its Applications, 38, 3, 547-551.

Phillips K. (1991), “A two-country model of stochastic output with changes in regime”, Journal of International Economics, 31, 121-142.
Politis D. N., and Romano J. P. (1994), “The stationary bootstrap”, *Journal of the American Statistical Association*, 89, 1303-1313.

Sims C. A., and Zha T. (2006), “Were there regime switches in U.S. monetary policy?”, *American Economic Review*, 96, 1, 54-81.

Taamouti A. (2012) “Moments of multivariate regime switching with application to risk-return trade-off”, *Journal of Empirical Finance*, 19, 292308.

Timmermann A. (2000), “Moments of Markov switching models”, *Journal of Econometrics*, 96, 75-111.

Yang M. (2000), “Some properties of vector autoregressive processes with Markov-switching coefficients”, *Econometric Theory*, 16, 23-43.