Moments given regimes in Markov-switching VAR models *

Gabriele Fiorentini\(^{(a)}\), Christophe Planas\(^{(b)}\) and Alessandro Rossi\(^{(b)}\)

March 2014

Abstract

We contribute to the theoretical understanding of Markov-switching Vector Auto-Regressive (MS VAR) processes by making available conditional moments given regimes — i.e. moments conditional on any state or sequence of states — up to the fourth order. These conditional moments have several utilities. They summarize the series level, variation and co-movements given the current state, hence enhancing model interpretation. They allow generalizing existing results about third and fourth order unconditional moments to the multivariate case. These moments can also be used to build a finite mixture of gaussian distributions that carefully approximates the marginal distribution of the MS VAR process either unconditionally or for a given regime. Our results also apply to the closely related MS state space models where some further applications are discussed.

JEL code: C1.

KEYWORDS: Markov-switching; VAR models; Higher-order moments; Mixtures of normals; State space models.

\*The views expressed in this paper are those of the authors and should not be attributed to the European Commission. We are grateful to Gaby Perez-Quiros and Enrique Sentana for their comments and suggestions. Special thanks are due to Massimo Guidolin for very helpful discussions. Of course, the usual caveat applies. Fiorentini acknowledges funding from MIUR PRIN MISURA - Multivariate models for risk assessment.\(\(a\)\) University of Florence, School of Economics, and Rimini Centre for Economic Analysis.\(\(b\)\) Joint Research Centre of the European Commission.

E-Mails: fiorentini@ds.unifi.it, christophe.planas@jrc.ec.europa.eu, alessandro.rossi@jrc.ec.europa.eu.
1 Introduction

Markov-switching (MS) 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 describe time series subject to changes in pattern. Outliers and structural breaks for instance can arise with random timing and magnitude. In macroeconomics, MS models have been introduced by Hamilton (1989) with the aim of capturing the asymmetry of the business cycle in the US GDP growth. Kim and Nelson (1999a) and McConnell and Perez-Quiros (2000) have later extended Hamilton’s model to account for the reduction in business cycle fluctuations known as the Great Moderation. Phillips (1991) has applied Hamilton’s model to a multi-country case. Ang and Beckaert (2002) underline the usefulness of the 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, 2008, 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 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 by Timmermann (2000), Yang (2000), and Francq and Zakoian (2001, 2002). These studies focus on stationarity issues and unconditional moments. We contribute to the statistical knowledge of MS VAR models by providing conditional moments given regimes — i.e. moments conditional on any state or sequence of states — up to the fourth order. The results for the first and second order moments are exact whereas higher-order moments are approximated up to any desired level of accuracy. These conditional moments have several utilities. First, they summarize the series level, variation and co-movements in the different regimes, hence helping the model interpre-
tation. In financial applications, co-skewness and co-kurtosis which capture the dependence between the level of one variable and the volatility or the extreme values of a reference variable in a given state are of particular interest (see Guidolin and Timmermann, 2008). Second, they allow generalizing Timmermann (2000) results about higher-order unconditional moments to the multivariate case. Third, the conditional means and covariances can be exploited to build a finite mixture of gaussian distributions that approximates marginal distributions of the MSVAR process, either unconditional or given regimes. Quantities of interest such as exceedance correlations, value-at-risk, or long-term predictive density can then be easily derived. In spite of their importance, so far the moments conditional on state realizations are available only in some special cases, for instance when the MS VAR model has no autoregressive terms. We derive results for the fully general case; Matlab and Fortran codes that implement the formulae given in the paper are available from the authors.

The general MS VAR framework is presented in Section 2 together with assumptions and notations. We focus on models with finite number of states and time-invariant transition probabilities. Section 3 gives our main results about conditional means and covariances. Section 4 considers higher-order moments, both unconditional and conditional on a state or a sequence of states. The accuracy of our results is analyzed through a case study where several configurations are examined. Section 5 shows how the marginal distribution of a MS VAR process can be approximated by an appropriate mixture of normal distributions whose weights are given by the probabilities attached to the different state sequences. Section 6 extends the results to the closely related MS state space models where some other applications are discussed. Section 7 concludes, while all proofs are gathered in the Appendix.

## 2 Model and assumptions

Let \((\Omega, \mathcal{F}, P)\) be a probability space on which a \(n\)-dimensional standard gaussian variable \(\{\varepsilon_t\}\) and an homogeneous \(K\)-state irreducible Markov chain \(\{S_t\}\) are defined at discrete time \(t\). The first-order MS VAR process is generated by the stochastic difference equation:

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

where \(x_t = (x_{1t}, \ldots, x_{nt})'\). The \(n \times 1\) vector \(\alpha_{S_t}\) and the \(n \times n\) matrices \(\Phi_{S_t}, \Lambda_{S_t}\) take \(K\) different values depending on the realization of the discrete latent variable \(S_t\). Specifications involving more lags can easily be cast into the formulation above through the VAR(1) companion form.
Throughout the paper we assume that model parameters are given; estimation procedures can be found in Hamilton (1989), Kim and Nelson (1999), and Fruhwirth-Schnatter (2006) for instance.

The joint process \{((\Phi_{S_t}, \alpha_{S_t}, \Lambda_{S_t}, \varepsilon_t), t \in \mathcal{N}\} inherits strict ergodic stationarity from \{(S_t, \varepsilon_t), t \in \mathcal{N}\}. We also assume that the top Lyapunov exponent associated to the process (2.1) is negative so the MS VAR model (2.1) is strictly stationary (see Brandt, 1986, Bougerol and Picard, 1992). As the initial condition \((\Lambda_{S_0}, \alpha_{S_0})\) belongs to \(L^r\) for any \(r > 0\), the existence of moments up to any order is guaranteed by Theorem 4.2 in Stelzer (2009).

We denote \(1_K\) the \(K \times 1\) vector with all elements equal to 1, \(e_\ell\) the \(K \times 1\) unit vector \(e_\ell = [0_{\ell-1}, 1, 0_{K-\ell}]^\prime\) for \(\ell = 1, \cdots, K\), \(I_n\) the \(n \times n\) identity matrix, and \(M\) the \(K \times K\) backward transition probability matrix with generic element \(m_{ji} = P(S_t = i | S_{t+1} = j)\). Backward and forward transition probabilities \(p_{ij} = P(S_{t+1} = j | S_t = i)\) are related through the equation \(m_{ji} \pi_j = p_{ij} \pi_i\), where \(\pi_\ell = P(S_\ell = k)\) represents the probability of being in state \(k\). For \(k = 1, \cdots, K\), we also denote \(J_k\) the \(n \times nK\) matrices \(J_k = [0_{n \times n(k-1)}, I_n, 0_{n \times n(K-k)}]\) which all together sum to \(J = \sum_{k=1}^K J_k\). Following Yang (2000), we define the \(K\)-block diagonalization operator for any \(n_1 K \times n_2\) matrix \(Q\) as:

\[
\text{diag}_K Q = \text{diag}_K \left[ \begin{array}{c} Q_1 \\ \vdots \\ Q_K \end{array} \right] = \left[ \begin{array}{ccc} Q_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Q_K \end{array} \right]_{n_1 K \times n_2 K}
\]

where the blocks \(Q_k\) are \(n_1 \times n_2\) matrices. The \(nK \times K\) matrix \(\alpha\) and the \(nK \times nK\) matrices \(\Lambda\) and \(\Phi\) are defined accordingly:

\[
\alpha = \text{diag}_K \left[ \begin{array}{c} \alpha_1 \\ \vdots \\ \alpha_K \end{array} \right], \quad \Lambda = \text{diag}_K \left[ \begin{array}{c} \Lambda_1 \\ \vdots \\ \Lambda_K \end{array} \right], \quad \text{and} \quad \Phi = \text{diag}_K \left[ \begin{array}{c} \Phi_1 \\ \vdots \\ \Phi_K \end{array} \right]
\]

Notice that the matrix \(J'\) inverts the \(\text{diag}_K\)-operator, i.e. for any \(nK \times n\) matrix \(Q\), \(\text{diag}_K Q J' = Q\). Lemma 1 below defines an operator which is useful to remove the blocks of zeroes when vectorizing a block-diagonal matrix.

**Lemma 1** Let \(A\) be a \(nK \times nK\) block-diagonal matrix made up of \(K\) blocks of dimension \(n \times n\). Then \(\text{vec}(A) = H \text{vec}(AJ')\) for the \(n^2 K^2 \times n^2 K\) matrix \(H\) such that \(H = \sum_{k=1}^K J'_k \otimes (J'_k J_k)\).

We now turn to the first and second conditional moments, i.e. the mean and variance of \(x_t\) conditional on any length-\(p\) sequence of states \(S_{t-p+1} \equiv (S_t, S_{t-1}, \cdots, S_{t-p+1}), p > 0\).
3 Conditional means and variances

We give our first results in the following theorem. Remind that the model parameters enter the conditioning sets but they are omitted for the sake of clarity.

**Theorem 1** The first and second moments of \( x_t \) conditionally on \( S^t_{t-p+1} \), \( p > 0 \), are given by:

(a). \( E(x_t \mid S^t_{t-p+1}) = \)

\[ = \alpha_s + \sum_{i=2}^{p} \prod_{j=2}^{i} \Phi_{s_{t-j+2}} \alpha_{s_{t-i+1}} + \prod_{j=1}^{p} \Phi_{s_{t-j+1}} J_{s_{t-p+1}} [I_{nK} - (M \otimes I_n) \Phi]^{-1} (M \otimes I_n) \alpha_1 \]

(b). \( E(x_t x'_t \mid S^t_{t-p+1}) = \)

\[ = \alpha_s \alpha'_s + I_{s_{t} - s_{t-1}} + \sum_{j=1}^{p-1} \prod_{i=1}^{j} \Phi_{s_{t-i+1}} (\alpha_{s_{t-j}} \alpha'_{s_{t-j}} + \Lambda_{s_{t-j}} \Lambda'_{s_{t-j}}) (\prod_{i=1}^{j} \Phi_{s_{t-i+1}})' \]

\[ + \prod_{i=1}^{p} \Phi_{s_{t-i+1}} J_{s_{t-p+1}} A J'_{s_{t-p+1}} (\prod_{i=1}^{p} \Phi_{s_{t-i+1}})' + \Gamma + \Gamma' \]

where the \( nK \times nK \) block-diagonal matrix \( A \) is such as:

\[ vec(A) = H \{ I_{n^2K} - [J \Phi \otimes ((M \otimes I_n) \Phi)]H \}^{-1} vec(A_0 J'), \]

and the \( n^2K^2 \times n^2K \) matrix \( H \) is given in Lemma 1 whereas the \( nK \times nK \) matrix \( A_0 \) is defined by:

\[ A_0 = diag_k \{ (M \otimes I_n)(\alpha \alpha' + \Lambda \Lambda').J' \} \]

The \( n \times n \) matrix \( \Gamma \) is equal to:

\[ = \sum_{k=1}^{p-1} \prod_{i=1}^{k} \Phi_{s_{t-i+1}} \alpha_{s_{t-k}} \alpha'_{s_{t-i}} + \sum_{j=1}^{p-2} \sum_{k=j+1}^{p} \prod_{i=1}^{k} \Phi_{s_{t-i+1}} \alpha_{s_{t-k}} \alpha'_{s_{t-i}} (\prod_{i=1}^{k} \Phi_{s_{t-i+1}})' \]

\[ + \prod_{i=1}^{p} \Phi_{s_{t-i+1}} J_{s_{t-p+1}} [I_{nK} - (M \otimes I_n) \Phi]^{-1} (M \otimes I_n) \alpha_1 \sum_{j=0}^{p-2} \alpha'_{s_{t-j}} (\prod_{i=1}^{j} \Phi_{s_{t-i+1}})' \]

\[ + \prod_{i=1}^{p} \Phi_{s_{t-i+1}} J_{s_{t-p+1}} B_0 \varepsilon'_{s_{t-p+1}} \alpha'_{s_{t-p+1}} (\prod_{i=1}^{p} \Phi_{s_{t-i+1}})' \]

\[ + \prod_{i=1}^{p} \Phi_{s_{t-i+1}} J_{s_{t-p+1}} B J'_{s_{t-p+1}} (\prod_{i=1}^{p} \Phi_{s_{t-i+1}})' \]

5
where the $nK \times K$ block-diagonal matrix $B_0$ is such that:

$$B_0 = \text{diag}_K \{ [I_{nK} - (M \otimes I_n)\Phi]^{-1}B_{01}1_K \} \tag{3.8}$$

with

$$B_{01} = \text{diag}_K \{ (M \otimes I_n)\alpha 1_K \} \tag{3.9}$$

and the $nK \times nK$ block-diagonal matrix $B$ is given by:

$$\text{vec}(B) = H \{ I_{n^2K} - [J\Phi \otimes ((M \otimes I_n)\Phi)]H \}^{-1}\text{vec}((M \otimes I_n)\Phi B_0\alpha'J') \tag{3.10}$$

All quantities involved in Theorem 1 are straightforwardly obtained from the model specification (2.1). For the sake of generality Theorem 1 considers the case of a sequence of states over $p$ periods. In applied works, moments given the concurrent state such as $E(x_t \mid S_t)$ and $V(x_t \mid S_t)$ are typically the ones of most interest, as illustrated in Example 1 below. Long sequences are relevant for instance for scenario analysis, but we will see that they are also helpful to calculate the higher-order moments as well as to approximate the marginal distribution of the MS VAR process.

As an alternative to the closed-form expressions above, the conditional moments can be estimated by simulation. Such an approach is computationally expensive, especially when the regime of interest is rare, and eventually can even be infeasible in large models.

Theorem 1 opens up a simple route for calculating the first two unconditional moments since:

$$E(x_t) = \sum_{j=1}^{K} P(S_t = j) E(x_t \mid S_t = j)$$

$$E(x_t x_t') = \sum_{j=1}^{K} P(S_t = j) E(x_t x_t' \mid S_t = j)$$

Yang (2000) and Francq and Zakoian (2001) offer alternative expressions for computing the first two unconditional moments in MS VAR models.

**Example 1**: Guidolin and Timmermann (GT, 2005) fit a MS VAR model to the monthly UK stocks and bonds excess returns together with the dividend yield over the period 1976-2 to 2000-12. They consider three regimes that they interpret as bear, normal, and bull market periods with steady state probabilities equal to 13%, 68%, and 19% respectively. Each regime impacts the intercept, the autoregressive matrix, and the variance-covariance matrix of the shocks. Table 1 shows the first two moments of the UK stocks and bonds excess returns conditional on
the different regimes as well as the unconditional ones, calculated with the parameter estimates given in Table 4 of GT’s paper.

The bear market is characterized by expected annualized returns that are strongly negative at -38% and -16% for stocks and bonds, and by a high volatility at 25% and 12% respectively. In this state the returns on the two assets are almost uncorrelated. Under the normal regime, the expected annual returns rise to about 19% and 10% for stocks and bonds, the volatility falls to about 12% and 9% respectively, while the correlation rises to 0.50. In the third state the expected annualized returns are about 3%, with greater volatility about 20% and 16%, and correlation at 0.58. The mean returns differ substantially across the three regimes whereas the volatility is lower in the normal market for both assets.

To further illustrate the moments under regime, Figure 1 shows the stocks and bonds series together with their respective mean conditional on the regime which is found to be the most likely in each time-period. The series vary around the lowest level during the bear market-periods, which occurs 13% of the sample dates. The two series vary around a higher level in the normal and bullish market periods, which occur in 68% and 19% of the sample dates respectively.

We now turn to the higher-order moments.

4 Higher-order moments

The conditional moments of order three and four, i.e. $E(x_{it}x_{jt}x_{kt} | S_{t-p+1}^t)$ and $E(x_{it}x_{jt}x_{kt}x_{lt} | S_{t-p+1}^t)$, $p > 0$, reveal some non-linear features of the process such as asymmetry and fat tails as well as some characteristics of the joint distribution such as co-skewness and co-kurtosis. They are thus most relevant in the context of non-linear MS VAR models. Deriving the higher-order moments analytically like in Theorem 2 is however intricate. Simple yet accurate approximations can instead be built by observing that the distribution of $x_t$ given $S_{t-p+1}^t$ tends to normality as the length $p$ of the conditioning sequence increases, i.e. $\lim_{p \to \infty} f(x_t | S_{t-p}^t) = \phi(x_t; \lim_{p \to \infty} E(x_t | S_{t-p}^t), \lim_{p \to \infty} V(x_t | S_{t-p}^t))$, where $\phi(\cdot; \mu, \Sigma)$ denotes the normal density with mean $\mu$ and variance $\Sigma$. The third and fourth conditional moments are thus virtually known for $p$ large enough. This property can be exploited to approximate the higher-order moments of $x_t$ given $S_{t-p+1}^t$ by augmenting the information set with $S_{t-q}^t$, $q \geq p$, and applying the law of iterated expectations as shown in the following theorem.
Theorem 2 Let $p$ and $q$ be two integers such that $q \geq p > 0$. (a) Conditionally on the sequence $S_{t-p+1}^t$, the third central moment associated to the triple $(x_{it}, x_{jt}, x_{kt})$ of $x_t$’s elements is such that:

$$E \left[ \prod_{\ell=i,j,k} (x_{\ell t} - E(x_{\ell t} \mid S_{t-p+1}^t)) \mid S_{t-p+1}^t \right] = \sum_{S_{t-q}^t} P(S_{t-q}^t \mid S_{t-p+1}^t)$$

$$\times \left\{ \sum_{(\ell_1, \ell_2, \ell_3, \ell_4) \in P_1} E \left[ \prod_{\ell=\ell_1, \ell_2} (x_{\ell t} - E(x_{\ell t} \mid S_{t-q}^t)) \mid S_{t-q}^t \right] (E(x_{\ell_3 t} \mid S_{t-q}^t) - E(x_{\ell_4 t} \mid S_{t-p+1}^t)) + \prod_{\ell=i,j,k} (E(x_{\ell t} \mid S_{t-q}^t) - E(x_{\ell t} \mid S_{t-p+1}^t)) \right\} + \sum_{S_{t-q}^t} P(S_{t-q}^t \mid S_{t-p+1}^t) \eta_{ijk}(S_{t-q}^t)$$

(4.1)

where $P_1 = \{(i, j, k), (i, k, j), (j, k, i)\}$, and $\eta_{ijk}(S_{t-q}^t)$ denotes the third central moment given $S_{t-q}^t$, $\eta_{ijk}(S_{t-q}^t) = E \left[ \prod_{\ell=i,j,k} (x_{\ell t} - E(x_{\ell t} \mid S_{t-q}^t)) \mid S_{t-q}^t \right]$, which verifies $\lim_{q \to \infty} \eta_{ijk}(S_{t-q}^t) = 0$.

(b) Conditionally on the sequence $S_{t-p+1}^t$, the fourth central moment associated to any possible quadruple $(x_{it}, x_{jt}, x_{kt}, x_{lt})$ is such that:

$$E \left[ \prod_{\ell=i,j,k,l} (x_{\ell t} - E(x_{\ell t} \mid S_{t-p+1}^t)) \mid S_{t-p+1}^t \right] = \sum_{S_{t-q}^t} P(S_{t-q}^t \mid S_{t-p+1}^t)$$

$$\times \left\{ \sum_{(\ell_1, \ell_2, \ell_3, \ell_4) \in P_2} E \left[ \prod_{\ell=\ell_1, \ell_2} (x_{\ell t} - E(x_{\ell t} \mid S_{t-q}^t)) \mid S_{t-q}^t \right] E \left[ \prod_{\ell=\ell_3, \ell_4} (x_{\ell t} - E(x_{\ell t} \mid S_{t-q}^t)) \mid S_{t-q}^t \right] \right\} + \sum_{(\ell_1, \ell_2, \ell_3, \ell_4) \in P_3} E \left[ \prod_{\ell=\ell_1, \ell_2} (x_{\ell t} - E(x_{\ell t} \mid S_{t-q}^t)) \mid S_{t-q}^t \right] \prod_{\ell=\ell_3, \ell_4} (E(x_{\ell t} \mid S_{t-q}^t) - E(x_{\ell t} \mid S_{t-p+1}^t)) + \prod_{\ell=i,j,k,l} (E(x_{\ell t} \mid S_{t-q}^t) - E(x_{\ell t} \mid S_{t-p+1}^t)) \right\} + \sum_{S_{t-q}^t} P(S_{t-q}^t \mid S_{t-p+1}^t) \kappa_{ijkl}(S_{t-q}^t) + \sum_{(\ell_1, \ell_2, \ell_3, \ell_4) \in Q} \eta_{\ell_1, \ell_2, \ell_3, \ell_4}(S_{t-q}^t) (E(x_{\ell_1 t} \mid S_{t-p+1}^t) - E(x_{\ell_1 t} \mid S_{t-q}^t))$$

(4.2)
where $\mathcal{P}_2 = \{(i,j,k,l), (i,k,j,l), (i,l,j,k)\}$, $\mathcal{P}_3 = \mathcal{P}_2 \cup \{(j,k,i,l), (j,l,i,k), (k,l,i,j)\}$, $\mathcal{Q} = \{(i,j,k,l), (i,j,l,k), (i,k,l,j), (j,k,l,i)\}$, $\eta_{ijk}(S^t_{l-q})$ is like in (4.1) above, $\kappa_{ijkl}(S^t_{l-q})$ is given by:

$$
\kappa_{ijkl}(S^t_{l-q}) = E \left[ \prod_{\ell=i,j,k,l} \left( x_{\ell t} - E(x_{\ell t} | S^t_{l-q}) \right) \right] S^t_{l-q} - \sum_{(\ell_1,\ell_2,\ell_3,\ell_4) \in \mathcal{P}_2} E \left[ \prod_{\ell=\ell_1,\ell_2} \left( x_{\ell t} - E(x_{\ell t} | S^t_{l-q}) \right) \right] E \left[ \prod_{\ell=\ell_3,\ell_4} \left( x_{\ell t} - E(x_{\ell t} | S^t_{l-q}) \right) \right] S^t_{l-q}
$$

and it verifies $\lim_{q \to \infty} \kappa_{ijkl}(S^t_{l-q}) = 0$.

All moments involved in formulae (4.1) and (4.2) are given in Theorem 1 except for $\eta_{ijk}(S^t_{l-q})$ and $\kappa_{ijkl}(S^t_{l-q})$. Both terms tend to zero as $q$ increases because the distribution $f(x_t | S^t_{l-q})$ tends to normality. Hence the expressions (4.1) and (4.2) can be used to approximate the third and the fourth central moment given $S^t_{l-1}$ by choosing $q \geq p$ and setting both $\eta_{ijk}(S^t_{l-q})$ and $\kappa_{ijkl}(S^t_{l-q})$ to their limit value of zero. For some standard specifications symmetry already holds given $S_t$, i.e. $\eta_{ijk}(S^t_{l-q}) = 0$ for $q \geq 0$: in such cases Theorem 2 yields the third conditional central moment exactly (see Example 2 below). Otherwise this approach yields an approximation error which converges to zero as $q$ increases. The following corollary gives similar formulae for the unconditional higher-order central moments.

**Corollary 1** Let $q$ be a non-negative integer. (a) The unconditional third central moment associated to the triple $(x_{it}, x_{jt}, x_{kt})$ of $x_t$’s elements is given by:

$$
E \left[ \prod_{\ell=i,j,k} \left( x_{\ell t} - E(x_{\ell t}) \right) \right] = \sum_{S^t_{l-q}} P(S^t_{l-q}) \times \left\{ \sum_{(\ell_1,\ell_2,\ell_3) \in \mathcal{P}_1} E \left[ \prod_{\ell=\ell_1,\ell_2} \left( x_{\ell t} - E(x_{\ell t} | S^t_{l-q}) \right) \right] S^t_{l-q} \right\} (E(x_{\ell_3 t} | S^t_{l-q}) - E(x_{\ell_3 t})) + \sum_{S^t_{l-q}} P(S^t_{l-q}) \eta_{ijk}(S^t_{l-q})
$$

(4.3)

where $\mathcal{P}_1$ and $\eta_{ijk}(S^t_{l-q})$ are defined in Theorem 2.

(b) The unconditional fourth central moment associated to the quadruple $(x_{it}, x_{jt}, x_{kt}, x_{lt})$ verifies:
\[
E \left[ \prod_{\ell=i,j,k,l} (x_{\ell t} - E(x_{\ell t})) \right] = \sum_{S_{t-q}^t} P(S_{t-q}^t) \times \\
\left\{ \sum_{(t_1, t_2, t_3, t_4) \in P_2} E \left[ \prod_{\ell=t_1, t_2} (x_{\ell t} - E(x_{\ell | S_{t-q}^t})) \right] S_{t-q}^t E \left[ \prod_{\ell=t_3, t_4} (x_{\ell t} - E(x_{\ell | S_{t-q}^t})) \right] S_{t-q}^t \right. \\
+ \sum_{(t_1, t_2, t_3, t_4) \in P_3} E \left[ \prod_{\ell=t_1, t_2} (x_{\ell t} - E(x_{\ell | S_{t-q}^t})) \right] S_{t-q}^t \prod_{\ell=t_3, t_4} (E(x_{\ell t} | S_{t-q}^t) - E(x_{\ell t})) + \\
+ \prod_{\ell=i,j,k,l} (E(x_{\ell t} | S_{t-q}^t) - E(x_{\ell t})) \right\} + \\
\left. \sum_{S_{t-q}^t} P(S_{t-q}^t) \left[ \kappa_{ijkl}(S_{t-q}^t) + \sum_{(t_1, t_2, t_3, t_4) \in Q} \eta_{t_1, t_2, t_3, t_4} (S_{t-q}^t) \left( E(x_{\ell t} | S_{t-q}^t) - E(x_{\ell t} | S_{t-q}^t) \right) \right] \right] \quad (4.4)
\]

where \( P_2, \ P_3, \ Q, \) and \( \kappa_{ijkl}(S_{t-q}^t) \) are defined in Theorem 2.

Like for the conditional moments, an approximation to the unconditional higher-order moments can be obtained by choosing \( q \geq 0 \) and setting both \( \eta_{ijkl}(S_{t-q}^t) \) and \( \kappa_{ijkl}(S_{t-q}^t) \) equal to zero in (4.3)-(4.4). This yields an approximation error which is equal to \( \sum_{S_{t-q}^t} P(S_{t-q}^t) \eta_{ijkl}(S_{t-q}^t) \) for the third-order moment and to \( \sum_{S_{t-q}^t} P(S_{t-q}^t) \left[ \kappa_{ijkl}(S_{t-q}^t) + \sum_{(t_1, t_2, t_3, t_4) \in Q} \eta_{t_1, t_2, t_3, t_4} (S_{t-q}^t) \times \right] \left( E(x_{\ell t} | S_{t-q}^t) - E(x_{\ell t} | S_{t-q}^t) \right) \) for the fourth one. Both errors vanish as \( q \) increases. Although our results apply more generally to multivariate models and allow the calculation of both conditional and unconditional higher-order moments, the accuracy achieved in approximating the unconditional higher-order moments can be evaluated using the formulae derived by Timmermann (2000) for univariate cases.

**Example 2:** Timmermann (2000) considers the following univariate Markov-switching autoregressive process:

\[
x_{t} - \mu_{Z_{t}} = \rho_{Z_{t-1}} (x_{t-1} - \mu_{Z_{t-1}}) + \sigma_{Z_{t}} \epsilon_{t} \quad (4.5)
\]

where \( Z_{t} \in \{1, 2\} \) is a two-state Markov chain. Model (4.5) is parameterized in terms of deviations from the mean \( E(x_{t} | Z_{t}) = \mu_{Z_{t}} \), contrary to (2.1) which is parameterized in terms
of intercepts, but it can be cast into the format (2.1) by defining \( S_t = (Z_t \times Z_{t-1}) \), \( \alpha_{S_t} = \mu_{Z_t} - \rho_{Z_{t-1}} \mu_{Z_{t-1}} \), \( \Phi_{S_t} = \rho_{Z_{t-1}} \), and \( \Lambda_{S_t} = \sigma_{Z_t} \). Since \( f(x_t|Z_t) \) is an infinite mixture of normals where all mixture-components share the same mean, the distribution of \( x_t \) is symmetric given \( Z_t \) and thus also given \( S_t \). Hence, for univariate or multivariate processes parameterized as in (4.5), using \( q = 0 \) and \( \eta_{ijk}(S_t) = 0 \) in Corollary 1 yields the unconditional skewness with no approximation error. We turn to the calculation of the coefficient of kurtosis.

When both \( \eta_{ijk}(S_t) \) and \( \kappa_{ijkl}(S_t) \) are null, Corollary 1 provides the exact kurtosis value. The models considered by Timmermann (2000) in figures 1-4 where the autoregressive coefficient \( \rho \) is equal to zero give important examples. We thus consider models with autoregressive dynamics. Timmermann (2000) screens the excess kurtosis that the process (4.5) can generate for parameter values such that \((\mu_1, \mu_2) = (1, -3)\), \((\sigma_1, \sigma_2) = (2, 4)\), \( \rho = 0.9 \) or \( (\rho_1, \rho_2) = (0.99, 0.81) \), and transition probability parameters that vary over the interval \((0, 1)\). We retain this setting but we restrict the transition probability parameters to \((p_{11}, p_{22}) \in \{(0.2, 0.2), (0.7, 0.7), (0.9, 0.9)\}\).

For the six model specifications, Figure 2 displays the exact kurtosis coefficient together with the value obtained by applying Corollary 1 with \( \kappa_{ijkl}(S^t_{t-q}) \) set equal to zero and \( q \) varying from 1 to 15.

Figure 2 shows that the convergence path depends on the process persistence. With transition probabilities as low as \((p_{11}, p_{22}) = (0.2, 0.2)\), the kurtosis is obtained almost exactly with very small values of \( q \) — see Figure 2 top panels. Convergence is slower when the latent Markov process is more persistent: for instance, when \( \rho = 0.9 \) and \( (p_{11}, p_{22}) = (0.9, 0.9) \), the target value is 3.448 while using Corollary 1 with \( q = 1, 5, \) and 10 yields 3.326, 3.426, and 3.446, respectively. In this case achieving an error below 1\% of the target value requires setting \( q \) equal to 5.

The persistence implied by the autoregressive roots also matters. When the autoregressive parameter switches between 0.99 and 0.81 and the transition probabilities are equal to \((0.7, 0.7)\), using Corollary 1 with \( q = 1, 5, \) and 10 yields a kurtosis coefficient equal to 3.026, 3.066, and 3.080, respectively, against an exact value at 3.085. In this case setting \( q \) equal to 4 is sufficient to get an approximation error below 1\% of the target value.

The most difficult situation arises when \((\rho_1, \rho_2) = (0.99, 0.81)\) and \((p_{11}, p_{22}) = (0.9, 0.9)\) as displayed in the bottom right panel of Figure 2. This configuration is rather extreme. In this situation setting \( q \) equal to 1, 5, and 15 yields a kurtosis coefficient equal to 3.060, 3.163, and 3.258 respectively, whereas the exact value is 3.290. Although the 5\% error threshold is already reached with \( q = 5 \), shrinking the error below 1\% requires increasing \( q \) until 15. This may
seem a rather large quantity but with our Fortran code this calculation requires only a couple of minutes. Increasing \( q \) further like for instance until 19 yields a kurtosis coefficient at 3.271, which is very close to the target value.

Overall this example suggests that the approximation to the higher-order unconditional moments obtained from Corollary 1 is reasonably accurate. In practice, one may start with a small value for \( q \) and increment it until the result remains almost unchanged. We illustrate the empirical relevance of the conditional and unconditional higher-order moments in the context of Example 1 introduced in Section 3.

**Example 1 cont’d** (GT, 2005): Table 2 below shows the skewness, kurtosis, co-skewness, and co-kurtosis, of the UK excess returns on stocks and bonds over the period 1976-1/2000-12, both unconditional and for a given regime, as implied by the parameter values given in Table 4 of GT’s paper. For any couple of variables \((x_{it}, x_{jt})\) the given-regime co-skewness and co-kurtosis are defined as:

\[
\text{CoSk}(x_{it}, x_{jt}|S_t) = \frac{E((x_{it} - E(x_{it}|S_t))(x_{jt} - E(x_{jt}|S_t))^2|S_t)}{\{V(x_{it}|S_t)V(x_{jt}|S_t)^2\}^{1/2}}
\]

\[
\text{CoK}_{13}(x_{it}, x_{jt}|S_t) = \frac{E((x_{it} - E(x_{it}|S_t))(x_{jt} - E(x_{jt}|S_t))^3|S_t)}{\{V(x_{it}|S_t)V(x_{jt}|S_t)^3\}^{1/2}}
\]

(4.6)

where \( V(\cdot) \) denotes the variance. For normally distributed variables, the co-skewness is null and the co-kurtosis above is equal to three times the correlation between \( x_{it} \) and \( x_{jt} \). According to Table 2, the unconditional marginal distributions of stocks and bonds are negatively skewed and leptokurtic. Under the bear market, the marginal distribution of stock returns remains close to normality but that of bonds is leptokurtic. Conversely, in the bull market, the marginal distribution of stocks returns is leptokurtic whereas that of bonds is close to normality. The joint distribution in the second state seems to be close to normality. Co-skewness is very low in all regimes, suggesting that returns on stocks (bonds) are not influenced by the volatility of bonds (stocks) returns whatever the market state. Co-kurtosis is more market-dependent: excess co-kurtosis is almost zero for the two variables in both bear and normal markets but it is high in the bull one. In this third regime the two variables take extreme values simultaneously with the same sign.
5 Approximating the marginal distribution via mixture of gaussians

The marginal distribution of MS VAR processes is generally unknown. Two exceptions occur when: (i) the model (2.1) has no autoregressive lag, i.e. $\Phi_{S_t} = 0$, in which case the marginal law of $x_t$ verifies $f(x_t) = \sum_{S_t} p(S_t) \phi(x_t; \alpha_{S_t}, \Lambda_{S_t}, \Lambda'_{S_t})$; and (ii) the model is parameterized in terms of conditional mean as in (4.5) but both the autoregressive coefficient and the shocks-loading matrix is time-invariant, i.e. $\Phi_{S_t} = \Phi$ and $\Lambda_{S_t} = \Lambda$, as detailed in the following example.

Example 3: Hamilton (1989) considers the univariate model:

$$x_t - \mu_{S_t} = \rho (x_{t-1} - \mu_{S_{t-1}}) + \sigma \epsilon_t \quad (5.1)$$

where $S_t$ is a two-state Markov process which determines the concurrent growth $\mu_{S_t}$. The condition $|\rho| < 1$ ensures both weak and strict stationarity. Iterating backward, the lagged terms $\mu_{S_{t-j}}, j > 0$, vanish so the infinite sum representation verifies:

$$x_t = \mu_{S_t} + \sum_{j=0}^{\infty} \rho^j \sigma \epsilon_{t-j}$$

Hence the conditional distribution $f(x_t|S_t)$ is gaussian and the marginal distribution of the observations can be written as:

$$f(x_t) = \sum_{S_t} p(S_t) \phi(x_t; \mu_{S_t}, \sigma^2/(1 - \rho^2))$$

The marginal distribution of $x_t$ in Hamilton’s switching growth model is thus exactly known. This implies the exact likelihood associated to model (5.1) is known and no pre-sample conditions are needed to evaluate it (see also Albert and Chib, 1993). It also implies that in Hamilton’s model, Corollary 1 with $q = 0$ yields the unconditional third and fourth central moments exactly. This result extends to any multivariate generalization of Hamilton’s model, like for instance in Phillips (1991).

Except for the two special cases (i) and (ii), the marginal distribution of $x_t$ is only known as the limit expression:

$$f(x_t) = \lim_{p \to \infty} \sum_{S_{t-p+1}^t} p(S_{t-p+1}^t) \phi \left( x_t; E(x_t|S_{t-p+1}^t), V(x_t|S_{t-p+1}^t) \right) \quad (5.2)$$
since the MS VAR model (2.1) is gaussian conditionally to any semi-infinite realization of the discrete latent variable. Because integrating over all possible semi-infinite regime paths is infeasible, the limit above is intractable. For any integer \( p > 0 \), we define \( f_p(\cdot) \) as the \( K^p \)-component mixture of normals as in:

\[
f_p(x_t) = \sum_{S_{t-p+1}} P(S_{t-p+1}) \phi \left( x_t; E(x_t|S_{t-p+1}), V(x_t|S_{t-p+1}) \right)
\]

where the mean and variance-covariance matrices of each mixture component are given in Section 3. By construction the distribution \( f_p(x_t) \) converges to the unknown marginal distribution of the MS VAR process as the number of mixture components increases. It can thus be used with finite values of \( p \) to approximate the unknown marginal distribution. The marginal distribution of \( x_t \) in a given regime \( S_t \) may also be of interest. It verifies the limit expression

\[
f(x_t|S_t) = \lim_{p \to \infty} \sum_{S_{t-p+1}} P(S_{t-p+1}|S_t) \phi \left( x_t; E(x_t|S_{t-p+1}), V(x_t|S_{t-p+1}) \right)
\]

Similarly to the unconditional case we build the approximating density \( f_p(x_t|S_t) \) such that:

\[
f_p(x_t|S_t) = \sum_{S_{t-p+1}} P(S_{t-p+1}|S_t) \phi \left( x_t; E(x_t|S_{t-p+1}), V(x_t|S_{t-p+1}) \right)
\]

which differs from (5.3) in the mixture weights and in the number of mixture components only.

In the next example, we compare the approximation (5.3) to a non-parametric estimate of the marginal density in an intricate situation where switches occur in growth, dynamics, and volatility.

**Example 4:** Let us consider the process:

\[
x_t = \begin{cases} 
1.5 + .5x_{t-1} + 1.5\epsilon_t & \text{if } S_t = 1 \\
-5.5 + .2x_{t-1} + .5\epsilon_t & \text{if } S_t = 2 
\end{cases}
\]

with \( P(S_t = i|S_{t-1} = i) = .9 \) for \( i = 1, 2 \). The top Lyapunov exponent is equal to .5 \( \log(1.5) + .5\log(.2) = -1.15 \), so the process (5.5) is strictly stationary. It can easily be checked that weak stationarity holds as well.

To estimate the marginal density, we generate \( G = 10^7 \) values for \( x_t \) from model (5.5) with starting condition set to \( x_{t-501} = 0 \) while \( S_{t-500} \) is drawn from its stationary distribution. The sample \( x_{t}^g, g = 1, 2, \ldots, G \), is used in the non-parametric density estimate:

\[
\hat{f}(x_t) = \frac{1}{Gh} \sum_{g=1}^{G} Ker \left( \frac{x_t - x_{t}^g}{h} \right)
\]
where $Ker(\cdot)$ is the gaussian kernel with smoothing factor $h$. Figure 3 shows the resulting marginal density estimates over the range $x_t \in (-10,10)$. As can be seen the distribution is bi-modal and asymmetric. Using Theorem 1 and equation (5.3) with $p = 1, 2, 3$, we obtain the following approximating mixtures:

$$f_1(x_t) = .5 \phi(x_t; 2.12, 5.17) + .5 \phi(x_t; -6.66, .57)$$

$$f_2(x_t) = .45 \phi(x_t; 2.56, 3.54) + .05 \phi(x_t; -1.83, 2.39) + .05 \phi(x_t; -5.08, .46) + .45 \phi(x_t; -6.83, .27)$$

$$f_3(x_t) = .405 \phi(x_t; 2.78, 3.13) + .045 \phi(x_t; .59, 2.85) + .005 \phi(x_t; -1.04, 2.36) + .045 \phi(x_t; -1.92, 2.32) + .045 \phi(x_t; -4.99, .39) + .005 \phi(x_t; -5.87, .35) + .045 \phi(x_t; -6.52, .27) + .045 \phi(x_t; -6.87, .26)$$

These three density estimates, which are obtained without computational cost, are shown in Figure 3. Although a discrepancy is visible, the two-component mixture already captures the main features of the kernel estimate. Increasing the number of components to four or eight makes the gaussian mixture indistinguishable from the kernel estimate. We could check that the eight-component mixture ($p = 3$) belongs to the 95% confidence interval around the non-parametric estimate (see Parzen, 1962).

The mixture approximation also applies to multivariate settings as illustrated in the next example.

Example 1 cont’d (GT 2005): Figure 4 shows the bivariate contours together with the marginal distributions of UK stocks and bonds monthly excess returns over the years 1976-2000, both given regime and unconditional. These distributions are estimated via the gaussian mixtures (5.3) and (5.4) with $p$ set equal to ten using the parameter values reported in Table 4 of GT’s paper.

As can be seen in Figure 4, the departure from normality of the unconditional joint distribution of stocks and bonds returns is mainly due to a thick negative tail on stocks returns. Under the bear regime, the marginal distribution of bonds excess returns is negatively skewed, in agreement with the moments displayed in Table 3. The intermediate state is the one where the joint distribution of stocks and bonds excess returns is closest to normality. Figure 4 also displays the modes of the joint distribution as well as the 1% and 99% quantiles. The vari-
ability of the 1% quantile across the three different states shows the worthiness of monitoring value-at-risk under each regime.

6 Markov-switching state space models

The MS VAR process (2.1) may also appear as transition equation of a Markov-switching state space (MS SS) model as in:

\[
\begin{align*}
y_t & = c_{S_t} + H_{S_t} x_t + \gamma_{S_t} u_t \\
x_t & = \alpha_{S_t} + \Phi_{S_t} x_{t-1} + \Lambda_{S_t} \epsilon_t
\end{align*}
\]  

(6.1)

where \( \epsilon_t \) and \( u_t \) are independent standard white noises. The vector \( y_t \) contains observations on the endogenous variables whereas the state \( x_t \) can be either partially or fully unobserved. The MS SS model (6.1) has been introduced in statistics by Harrisson and Stevens (1976). Applications and exhaustive discussions can be found in Shumway and Stoffer (1991) as well as in the textbooks by Fruhwirth-Schnatter (2006, Chapter 13) and Kim and Nelson (1999b). Theorem 1-2 extend readily to MS SS models, with the first two moments being such as:

\[
\begin{align*}
E(y_t|S_{t-p+1}^t) & = c_{S_t} + H_{S_t} E(x_t|S_{t-p+1}^t) \\
V(y_t|S_{t-p+1}^t) & = H_{S_t} V(x_t|S_{t-p+1}^t) H_{S_t} + \gamma_{S_t} \gamma^t_{S_t}
\end{align*}
\]  

(6.2)

These moments can be plugged into formulae (4.1) and (4.2) to calculate the (co-)skewness and (co-)kurtosis of the vector \( y_t \). We illustrate below the usefulness of the higher-order moments for revealing nonlinearities in a multivariate state space model.

Example 5: Chauvet (1998) and Kim and Nelson (1998) consider a Markov-switching dynamic factor model in order to extract a composite index of the US business cycle from the growth rates of the US industrial production index, the non-farm payroll employment, the personal income less transfer payments, and the real manufacturing and trade sales. Their dynamic factor model is specified as in:

\[
\begin{align*}
y_{it} & = \lambda_i f_t + v_{it} \\
f_{t} & = \mu_{S_t} + a_t \\
v_{it} & = \phi_{i1} v_{it-1} + \phi_{i2} v_{it-2} + \sigma_i \epsilon_{it}
\end{align*}
\]  

(6.3)

where \( a_t, \epsilon_{it}, i = 1, \cdots, 4 \), 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 $S_t \in \{1, 2\}$. Model (6.3) is easily cast into the MS SS format (6.1). Camacho, Perez-Quiros, and Poncela (CPP, 2012) estimate model (6.3) on monthly observations from January 1967 to November 2010.

The higher-order moments are helpful to analyze some non-linear aspects of the model fit. We consider the co-skewness defined in (4.6) and the co-kurtosis statistics $CoK_{22}(y_{it}, y_{jt})$ defined as:

$$CoK_{22}(y_{it}, y_{jt}) = \frac{E((y_{it} - E(y_{it}))^2)(y_{jt} - E(y_{jt}))^2)}{V(y_{it})V(y_{jt})}$$  (6.4)

If $y_{it}, y_{jt}$ were jointly normally distributed with correlation coefficient $\rho$, then the co-kurtosis $CoK_{22}$ would be equal to $1 + 2\rho^2$; larger values indicate a propensity of large shocks to occur simultaneously. Table 3 shows the univariate skewness, kurtosis, as well as the co-skewness (4.6) and co-kurtosis for the growth rate of Industrial Production (IP) and Employment (E). The empirical univariate higher-order moments suggests that the two variables are subject to large shocks which are predominantly negative. The negative value taken by the co-skewness between Employment and Industrial Production also shows that the growth of Employment tends to be below average when Industrial Production goes through a period of large volatility. Finally, the large co-kurtosis indicates co-movements in volatility.

Table 3 also shows the higher-order moments yielded by the model. To compute the model-based moments we use the parameter estimates given in Table 3 of CPP’s paper. The conditional distribution $f(y_t|S_t)$ in (6.3) is gaussian so setting equal to one in Corollary 1 gives all higher-order moments exactly. As can be seen in Table 3, model (6.3) captures the asymmetry of Employment and Industrial Production series. The model however under-weights the frequency of outlier occurrences and it does not foresee co-movements in volatility. Model (6.3) could thus be upgraded by letting the variance of the common shock $a_t$ switching between a normal regime and a regime of large volatility.

In the MS SS framework, Theorem 1 has further applications. For instance Kim (1994) has devised an algorithm for classical inference on model parameters and unobserved quantities for such models. To initialize his algorithm, Kim treats the pre-sample quantities $x_0$ and $x_{-1}$ as additional parameters to be estimated. Alternatively, use can be made of Theorem 1 to launch Kim’s filter directly from $E(x_1|S_1, S_0)$ and $V(x_1|S_1, S_0)$. Kim’s filter initialization also requires the probability $P(S_0, S_1|y_1) \propto f(y_1|S_0, S_1)P(S_0, S_1)$, but $f(y_1|S_0, S_1)$ is generally unknown. An approximation to $f(y_1|S_0, S_1)$ can however be derived from $f_q(x_1|S_0, S_1)$ since, if $x_1$ can be
approximated by a mixture of normals, then $y_1$ in (6.1) admits a similar approximation:

$$f_q(y_1|S_0, S_1) = \sum_{S_{-1}, \ldots, S_{-q}} P(S_{-1}, \ldots, S_{-q}|S_0) \phi(y_1; E(y_1|S_1, \ldots, S_{-q}), V(y_1|S_1, \ldots, S_{-q}))$$

(6.5)

where the conditional moments are as in (6.2).

Finally, the results in this paper are also relevant for the Bayesian analysis of MS SS models. First, when evaluating the conditional likelihood $f(y_1, \ldots, y_t|S_t)$, which comes into play in Markov Chain Monte Carlo sampling, the Kalman filter can be exactly initialized using $E(x_1|S_1)$ and $V(x_1|S_1)$ as given in Theorem 1 (see Fiorentini, Planas an Rossi, 2014). Next, sampling the discrete latent variable is best performed using the efficient algorithm proposed by Gerlach, Carter, and Kohn (2000). Yet, as $f(y_1|S_1)$ is generally an unknown mixture, the starting point $S_1$ cannot be sampled without some simplifying assumption. The previous results make possible using a gaussian sum approximation to the true distribution $f(y_1|S_1)$ similarly to (6.5).

7 Conclusion

We make available the moments of MS VAR processes conditional on the current regime up to the fourth order. These conditional moments enrich the model interpretation by summarizing the model properties in the different regimes. They also yield the model-based unconditional skewness and kurtosis, two quantities that were so far unavailable in multivariate applications. The first two conditional moments can be used to build a careful approximation to the marginal distribution of the MS VAR process. Both the unconditional marginal distribution and the marginal distribution in a given regime can be obtained. This yields a detailed picture of the data characteristics as inferred by the MS VAR model at hand. Quantities of interest such as exceedance correlations, value-at-risk as well as the long-term predictive density are then easily obtained. The results in this paper apply to the closely related MS SS models where some further utilities arise both in the classical and in the Bayesian framework.
References


Statistics, 81, 608-616.


Appendix

Proof of Lemma 1 For any $nK \times n$ matrix $Q$, the $\text{diag}_K$ operator verifies $\text{diag}_K Q = \sum_{k=1}^K J'_k J_k Q J_k$. Hence $A = \text{diag}_K \{AJ\}' = \sum_{k=1}^K J'_k J_k A J'_k J_k$ and thus $\text{vec}(A) = \sum_{k=1}^K \text{vec}(J'_k J_k A J'_k) = \sum_{k=1}^K [J'_k \otimes (J'_k J_k)] \text{vec}(AJ')$. 

Proof of Theorem 1 (a) Conditionally on current and past history of the discrete latent variable, model (2.1) can be expressed as:

$$x_t = \sum_{j=0}^{\infty} \Phi(t, j)(\alpha S_{t-j} + \Lambda S_{t-j} \epsilon_{t-j})$$

(A.1)

where

$$\begin{align*}
\Phi(t, j) &= I_n & \text{if } j = 0 \\
\Phi(t, j) &= \Phi S_t \Phi S_{t-1} \cdots \Phi S_{t-j+1} & \text{if } j > 0
\end{align*}$$

Taking expectation conditional on $S_{t-p+1}$ yields:

$$E(x_t|S_{t-p+1}) = \alpha S_t + \sum_{i=2}^{p} \prod_{j=2}^{i} \Phi S_{t-j+2} \alpha S_{t-i+1} + \sum_{j=p}^{\infty} E[\Phi(t, j)\alpha S_{t-j}|S_{t-p+1}]$$

since $E(\epsilon_{t-j}|S_{t-p+1}) = 0$. To solve the infinite sum above we proceed as follows.

For $j = p$:

$$E[\Phi(t, p)\alpha S_{t-p}|S_{t-p+1}] = \prod_{i=1}^{p} \Phi S_{t-i+1} \sum_{S_{t-p}} \alpha S_{t-p} m(S_{t-p}, S_{t-p+1}) = \prod_{i=1}^{p} \Phi S_{t-i+1} J_{S_{t-p+1}} (M \otimes I_n) \alpha 1_K$$

For $j = p + 1$:

$$E[\Phi(t, p+1)\alpha S_{t-p-1}|S_{t-p+1}] = \prod_{i=1}^{p} \Phi S_{t-i+1} \sum_{S_{t-p} S_{t-p-1}} \sum_{S_{t-p-1}} \Phi S_{t-p} \alpha S_{t-p-1} m(S_{t-p-1}, S_{t-p}) m(S_{t-p}, S_{t-p+1}) = \prod_{i=1}^{p} \Phi S_{t-i+1} J_{S_{t-p+1}} (M \otimes I_n) \Phi (M \otimes I_n) \alpha 1_K$$

For $j > p$ the general term is such as:
$$E[\Phi(t, j)|S_{t-p+1}^t] = \prod_{i=1}^{p} \Phi_{S_{t-i+1}} J_{S_{t-p+1}} [(M \otimes I_n)\Phi]^{j-p} (M \otimes I_n)\alpha 1_K$$

Hence, under stationarity, the infinite sum verifies:

$$\sum_{j=p}^{\infty} E[\Phi(t, j)|S_{t-p+1}^t] = \prod_{j=1}^{p} \Phi_{S_{t-j+1}} J_{S_{t-p+1}} [I_{nK} - (M \otimes I_n)\Phi]^{-1} (M \otimes I_n)\alpha 1_K \quad (A.2)$$

which proves result (a) in Theorem 1.

**Proof of Theorem 1 (b)** Expression (A.1) also implies:

$$E(x_t x'_t|S_{t-p+1}^t) = \sum_{j=0}^{\infty} E \left[ \Phi(t, j)(\alpha_{S_{t-j}} + \Lambda_{S_{t-j}} \epsilon_{t-j})(\alpha_{S_{t-j}} + \Lambda_{S_{t-j}} \epsilon_{t-j})' \Phi(t, j)'|S_{t-p+1}^t \right]$$

$$+ \sum_{j=0}^{\infty} \sum_{k=j+1}^{\infty} E \left[ \Phi(t, k)(\alpha_{S_{t-k}} + \Lambda_{S_{t-k}} \epsilon_{t-k})(\alpha_{S_{t-j}} + \Lambda_{S_{t-j}} \epsilon_{t-j})' \Phi(t, j)'|S_{t-p+1}^t \right]$$

$$+ \sum_{j=0}^{\infty} \sum_{k=j+1}^{\infty} E \left[ \Phi(t, j)(\alpha_{S_{t-j}} + \Lambda_{S_{t-j}} \epsilon_{t-j})(\alpha_{S_{t-k}} + \Lambda_{S_{t-k}} \epsilon_{t-k})' \Phi(t, k)'|S_{t-p+1}^t \right]$$

We treat the terms of the right hand side above separately.

(i) \( \sum_{j=0}^{\infty} E \left[ \Phi(t, j)(\alpha_{S_{t-j}} + \Lambda_{S_{t-j}} \epsilon_{t-j})(\alpha_{S_{t-j}} + \Lambda_{S_{t-j}} \epsilon_{t-j})' \Phi(t, j)'|S_{t-p+1}^t \right] \):

The sum of the first \( p \) terms verifies:

$$\sum_{j=0}^{p-1} E \left[ \Phi(t, j)(\alpha_{S_{t-j}} + \Lambda_{S_{t-j}} \epsilon_{t-j})(\alpha_{S_{t-j}} + \Lambda_{S_{t-j}} \epsilon_{t-j})' \Phi(t, j)'|S_{t-p+1}^t \right] =$$

$$= \alpha_{S_{t}} \alpha'_{S_{t}} + \Lambda_{S_{t}} \Lambda'_{S_{t}} + \sum_{j=1}^{p-1} \prod_{i=1}^{j} \Phi_{S_{t-i+1}} (\alpha_{S_{t-j}} \alpha'_{S_{t-j}} + \Lambda_{S_{t-j}} \Lambda'_{S_{t-j}}) (\prod_{i=1}^{j} \Phi_{S_{t-i+1}})'$$

which corresponds to the first row of (3.1).
For \( j = p \), the \((p + 1)\)-th term of the sum is such as:

\[
E \left[ \Phi(t, p)(\alpha_{s_t-p} + \Lambda_{s_t-p} \epsilon_{t-p})(\alpha_{s_t-p} + \Lambda_{s_t-p} \epsilon_{t-p})' \Phi(t, p)' \bigg| S_{t-p+1}^t \right] = 
\]

\[
= \prod_{i=1}^{p} \Phi_{s_{t-i+1}} \left[ \sum_{s_{t-p}} (\alpha_{s_{t-p}} \alpha_{s_{t-p}}' + \Lambda_{s_{t-p}} \Lambda_{s_{t-p}}') m(s_{t-p}, S_{t-p+1}) \right] \prod_{i=1}^{p} \Phi_{s_{t-i+1}}' 
\]

\[
= \prod_{i=1}^{p} \Phi_{s_{t-i+1}} J_{s_{t-i+1}} (M \otimes I_n)(\alpha \alpha' + \Lambda \Lambda') J' \left( \prod_{i=1}^{p} \Phi_{s_{t-i+1}} \right)'
\]

The last equality results from the definition of \( A_0 \) given in (3.3).

For \( j = p + 1 \):

\[
E \left[ \Phi(t, p + 1)(\alpha_{s_{t-p-1}} + \Lambda_{s_{t-p-1}} \epsilon_{t-p-1})(\alpha_{s_{t-p-1}} + \Lambda_{s_{t-p-1}} \epsilon_{t-p-1})' \Phi(t, p + 1)' \bigg| S_{t-p+1}^t \right] = 
\]

\[
= \prod_{i=1}^{p} \Phi_{s_{t-i+1}} \sum_{s_{t-p}} \Phi_{s_{t-p}} \left[ \sum_{s_{t-p-1}} (\alpha_{s_{t-p-1}} \alpha_{s_{t-p-1}}' + \Lambda_{s_{t-p-1}} \Lambda_{s_{t-p-1}}') m(s_{t-p-1}, S_{t-p}) \right] \Phi_{s_{t-p}}' 
\]

\[
\times m(s_{t-p}, S_{t-p+1}) \left( \prod_{i=1}^{p} \Phi_{s_{t-i+1}} \right)'
\]

\[
= \prod_{i=1}^{p} \Phi_{s_{t-i+1}} J_{s_{t-i+1}} (M \otimes I_n) \Phi_{s_{t-p}} \Phi_{s_{t-p}}' \left( \prod_{i=1}^{p} \Phi_{s_{t-i+1}} \right)'
\]

Defining the \( nK \times nK \) block-diagonal matrix \( A_1 \) such as \( A_1 = \text{diag}_K \{(M \otimes I_n) \Phi_{A_0} \Phi' J'\} \), the last equation above can be written as:

\[
E \left[ \Phi(t, p + 1)(\alpha_{s_{t-p-1}} + \Lambda_{s_{t-p-1}} \epsilon_{t-p-1})(\alpha_{s_{t-p-1}} + \Lambda_{s_{t-p-1}} \epsilon_{t-p-1})' \Phi(t, p + 1)' \bigg| S_{t-p+1}^t \right] = 
\]

\[
= \prod_{i=1}^{p} \Phi_{s_{t-i+1}} J_{s_{t-i+1}} A_{1} J'_{s_{t-i+1}} \left( \prod_{i=1}^{p} \Phi_{s_{t-i+1}} \right)'
\]

For \( j > p \), the general term verifies:

\[
E \left[ \Phi(t, j)(\alpha_{s_{t-j}} + \Lambda_{s_{t-j}} \epsilon_{t-j})(\alpha_{s_{t-j}} + \Lambda_{s_{t-j}} \epsilon_{t-j})' \Phi(t, j)' \bigg| S_{t-p+1}^t \right] = 
\]

\[
= \prod_{i=1}^{p} \Phi_{s_{t-i+1}} J_{s_{t-i+1}} A_{j-p} J'_{s_{t-i+1}} \left( \prod_{i=1}^{p} \Phi_{s_{t-i+1}} \right)'
\]

24
where the $nK \times nK$ block-diagonal matrix $A_{j-p}$ satisfies the recursion $A_{j-p} = \text{diag}_K \{(M \otimes I_n)\Phi A_{j-p-1}\Phi'J'\}$. Hence for $j \geq p$ the sum verifies:

$$
\sum_{j=p}^{\infty} E \left[ \Phi(t, j)(\alpha_{S_{t-j}} + \Lambda_{S_{t-j}}\epsilon_{t-j}) (\alpha_{S_{t-j}} + \Lambda_{S_{t-j}}\epsilon_{t-j})' \Phi(t, j)' \big| S_{t-p+1}^j \right] =
$$

$$
= \prod_{i=1}^{p} \Phi_{S_{t-i+1}} J_{S_{t-p+1}} \sum_{j=p}^{\infty} A_{j-p} J_{S_{t-p+1}}' \prod_{i=1}^{p} \Phi_{S_{t-i+1}}'
$$

To solve for $A = \sum_{j=p}^{\infty} A_{j-p}$ we observe that:

$$
J_{S_{t-p+1}} A_{0} J_{S_{t-p+1}}' = J_{S_{t-p+1}} (M \otimes I_n) (\alpha \alpha' + \Lambda \Lambda') J'
$$

$$
J_{S_{t-p+1}} A_{1} J_{S_{t-p+1}}' = J_{S_{t-p+1}} (M \otimes I_n) \Phi A_{0} \Phi' J'
$$

$$
\vdots
$$

$$
J_{S_{t-p+1}} A_{j-p} J_{S_{t-p+1}}' = J_{S_{t-p+1}} (M \otimes I_n) \Phi A_{j-p-1} \Phi' J'
$$

$$
\vdots
$$

Hence for $S_{t-p+1} = 1, \ldots, K$ the infinite sum $A$ verifies:

$$
J_A J_1 = J_1 (M \otimes I_n) (\alpha \alpha' + \Lambda \Lambda') J' + J_1 (M \otimes I_n) \Phi A \Phi' J'
$$

$$
\vdots
$$

$$
J_K A J_K' = J_K (M \otimes I_n) (\alpha \alpha' + \Lambda \Lambda') J' + J_K (M \otimes I_n) \Phi A \Phi' J'
$$

Using the definition of $A_0$ in (3.3) the system above can be written as:

$$
A J' = A_0 J' + (M \otimes I_n) \Phi A \Phi' J'
$$

Since $\text{vec}(A) = H \text{vec}(AJ')$ where $H$ is defined in Lemma 1, we can easily obtain (3.2).
(ii) To find $\Gamma$ we split the double-sum in three components:

$$
\Gamma = \sum_{j=0}^{\infty} \sum_{k=j+1}^{\infty} E \left[ \Phi(t, k)(\alpha_{S_{t-k}} + \Lambda_{S_{t-k}} \epsilon_{t-k})(\alpha_{S_{t-j}} + \Lambda_{S_{t-j}} \epsilon_{t-j})' \Phi(t, j)' | S_{t-p+1}^t \right]
$$

$$
= \sum_{j=0}^{p-2} \sum_{k=j+1}^{p-1} E \left[ \Phi(t, k)\alpha_{S_{t-k}} \alpha_{S_{t-j}}' \Phi(t, j)' | S_{t-p+1}^t \right] +
$$

$$
+ \sum_{j=0}^{p-2} \sum_{k=p}^{\infty} E \left[ \Phi(t, k)\alpha_{S_{t-k}} \alpha_{S_{t-j}}' \Phi(t, j)' | S_{t-p+1}^t \right] +
$$

$$
+ \sum_{j=p-1}^{\infty} \sum_{k=j+1}^{\infty} E \left[ \Phi(t, k)\alpha_{S_{t-k}} \alpha_{S_{t-j}}' \Phi(t, j)' | S_{t-p+1}^t \right]
$$

Given $S_{t-p+1}^t$ the first sum in the equation above verifies:

$$
\sum_{j=0}^{p-2} \sum_{k=j+1}^{p-1} E \left[ \Phi(t, k)\alpha_{S_{t-k}} \alpha_{S_{t-j}}' \Phi(t, j)' | S_{t-p+1}^t \right] =
$$

$$
= \prod_{i=1}^{p} \Phi_{S_{t-i+1}} \alpha_{S_{t-i}} + \sum_{j=1}^{p-2} \sum_{k=j+1}^{p-1} \prod_{i=1}^{k} \Phi_{S_{t-i+1}} \alpha_{S_{t-i}} \alpha_{S_{t-j}}' (\prod_{i=1}^{k} \Phi_{S_{t-i+1}})'
$$

which gives (3.4). The second sum can be solved using (A.2) as in:

$$
\sum_{j=0}^{p-2} \sum_{k=p}^{\infty} E \left[ \Phi(t, k)\alpha_{S_{t-k}} \alpha_{S_{t-j}}' \Phi(t, j)' | S_{t-p+1}^t \right] =
$$

$$
= \prod_{i=1}^{p} \Phi_{S_{t-i+1}} J_{S_{t-p+1}} [I_n - (M \otimes I_n)\Phi]^{-1} (M \otimes I_n)\alpha_{S_{t-i+1}} \prod_{i=1}^{p} \Phi_{S_{t-i+1}} \alpha_{S_{t-j}}' (\prod_{i=1}^{j} \Phi_{S_{t-i+1}})'
$$

which gives (3.5).

To calculate the last sum, we first focus on $j = p - 1$.

For $j = p - 1$ and $k = p$:

$$
E \left[ \Phi(t, p)\alpha_{S_{t-p}} \alpha_{S_{t-p+1}}' \Phi(t, p-1)' | S_{t-p+1}^t \right] = \prod_{i=1}^{p} \Phi_{S_{t-i+1}} J_{S_{t-p+1}} (M \otimes I_n)\alpha_{S_{t-p+1}} \prod_{i=1}^{p} \Phi_{S_{t-i+1}} \alpha_{S_{t-p+1}}' (\prod_{i=1}^{p} \Phi_{S_{t-i+1}})'
$$

$$
= \prod_{i=1}^{p} \Phi_{S_{t-i+1}} J_{S_{t-p+1}} B_{01} e_{S_{t-p+1}} \alpha_{S_{t-p+1}}' (\prod_{i=1}^{p} \Phi_{S_{t-i+1}})'
$$

26
where \( B_{01} = (M \otimes I_n)\alpha 1_K \) as in (3.9).

For \( j = p-1, k = p+1 \):

\[
E \left[ \Phi(t, p+1)\alpha_{S_{t-p+1}} \alpha'_{S_{t-p+1}} \Phi(t, p-1)' \big| S_{t-p+1}^t \right] =
\]

\[
= \prod_{i=1}^p \Phi_{S_{t-i+1}} J_{S_{t-p+1}} (M \otimes I_n) \Phi B_{01} 1_K \alpha'_{S_{t-p+1}} \left( \prod_{i=1}^{p-1} \Phi_{S_{t-i+1}} \right)' =
\]

\[
= \prod_{i=1}^p \Phi_{S_{t-i+1}} J_{S_{t-p+1}} B_{02} e_{S_{t-p+1}} \alpha'_{S_{t-p+1}} \left( \prod_{i=1}^{p-1} \Phi_{S_{t-i+1}} \right)' =
\]

where \( B_{02} \) is the \( nK \times K \) block-diagonal matrix such that \( B_{02} = diag_K \{ (M \otimes I_n) \Phi B_{01} 1_K \} \).

For \( j = p-1 \), the general term for \( k > p+1 \) is such as:

\[
E \left[ \Phi(t, k)\alpha_{S_{t-k}} \alpha'_{S_{t-p+1}} \Phi(t, p-1)' \big| S_{t-p+1}^t \right] =
\]

\[
= \prod_{i=1}^p \Phi_{S_{t-i+1}} J_{S_{t-p+1}} (M \otimes I_n) \Phi B_{0k-p+1} 1_K \alpha'_{S_{t-p+1}} \left( \prod_{i=1}^{p-1} \Phi_{S_{t-i+1}} \right)' =
\]

\[
= \prod_{i=1}^p \Phi_{S_{t-i+1}} J_{S_{t-p+1}} B_{0k-p+1} e_{S_{t-p+1}} \alpha'_{S_{t-p+1}} \left( \prod_{i=1}^{p-1} \Phi_{S_{t-i+1}} \right)' =
\]

where \( B_{0k-p+1} \) is the \( nK \times K \) block-diagonal matrix which satisfies the recursion \( B_{0k-p+1} = diag_K \{ (M \otimes I_n) \Phi B_{0k-p+1} 1_K \} \). Hence, for \( j = p-1 \) we have:

\[
\sum_{k=p} B_{0k-p+1} =
\]

\[
= \prod_{i=1}^p \Phi_{S_{t-i+1}} J_{S_{t-p+1}} \sum_{k=p}^\infty B_{0k-p+1} e_{S_{t-p+1}} \alpha'_{S_{t-p+1}} \left( \prod_{i=1}^{p-1} \Phi_{S_{t-i+1}} \right)' =
\]

The sum \( B_0 = \sum_{k=p}^\infty B_{0k-p+1} \) satisfies the system:

\[
J_1 B_0 J_1' = J_1 (M \otimes I_n) \alpha 1_K + J_1 (M \otimes I_n) \Phi B_{01} 1_K
\]

\[
\vdots
\]

\[
J_K B_0 J_K' = J_K (M \otimes I_n) \alpha 1_K + J_K (M \otimes I_n) \Phi B_{01} 1_K
\]

Using the definition of \( B_{01} \) in (3.9) the system above can be written as:

\[
B_0 1_K = B_{01} 1_K + (M \otimes I_n) \Phi B_{01} 1_K
\]

27
which gives $B_0$ as in (3.8). This proves the term in (3.6).

Next we increment $j$ to $j = p$ and focus on $\sum_{k=p+1}^{\infty} E \left[ \Phi(t, k) \alpha_{S_{t-p}'} \alpha_{S_{t-p+1}}' \Phi(t, p)|S^t_{t-p+1} \right]$.

For $j = p$ and $k = p + 1$:

$$E \left[ \Phi(t, p + 1) \alpha_{S_{t-p-1}'} \alpha_{S_{t-p+1}}' \Phi(t, p)|S^t_{t-p+1} \right] =$$

$$= \prod_{i=1}^{p} \Phi_{S_{t-i+1}'} J_{S_{t-p+1}} (M \otimes I_n) \Phi B_{01} \alpha^' J' (\prod_{i=1}^{p} \Phi_{S_{t-i+1}})'$$

$$= \prod_{i=1}^{p} \Phi_{S_{t-i+1}'} J_{S_{t-p+1}} B_{12} J'_{S_{t-p+1}} (\prod_{i=1}^{p} \Phi_{S_{t-i+1}})'$$

where the $nK \times nK$ matrix $B_{12}$ verifies $B_{12} = diag_K \{(M \otimes I_n) \Phi B_{01} \alpha^' J' \}$.

The next term for $j = p$ and $k = p + 2$ is such as:

$$E \left[ \Phi(t, p + 2) \alpha_{S_{t-p-2}'} \alpha_{S_{t-p+1}}' \Phi(t, p)|S^t_{t-p+1} \right] =$$

$$= \prod_{i=1}^{p} \Phi_{S_{t-i+1}'} J_{S_{t-p+1}} (M \otimes I_n) \Phi B_{02} \alpha^' J' (\prod_{i=1}^{p} \Phi_{S_{t-i+1}})'$$

$$= \prod_{i=1}^{p} \Phi_{S_{t-i+1}'} J_{S_{t-p+1}} B_{13} J'_{S_{t-p+1}} (\prod_{i=1}^{p} \Phi_{S_{t-i+1}})'$$

where the $nK \times nK$ matrix $B_{13}$ verifies $B_{13} = diag_K \{(M \otimes I_n) \Phi B_{02} \alpha^' J' \}$.

For $j = p$, the general term for $k > p + 2$ is given by:

$$E \left[ \Phi(t, k) \alpha_{S_{t-k}'} \alpha_{S_{t-p+1}}' \Phi(t, p)|S^t_{t-p+1} \right] =$$

$$= \prod_{i=1}^{p} \Phi_{S_{t-i+1}'} J_{S_{t-p+1}} (M \otimes I_n) \Phi B_{0k-p} \alpha^' J' (\prod_{i=1}^{p} \Phi_{S_{t-i+1}})'$$

$$= \prod_{i=1}^{p} \Phi_{S_{t-i+1}'} J_{S_{t-p+1}} B_{1k-p+1} J'_{S_{t-p+1}} (\prod_{i=1}^{p} \Phi_{S_{t-i+1}})'$$

where the $nK \times nK$ matrix $B_{1k-p+1}$ verifies the recursion $B_{1k-p+1} = diag_K \{(M \otimes I_n) \Phi B_{0k-p} \alpha^' J' \}$.

Hence, for $j = p$, summing over $k = p + 1, \ldots$ yields:

$$\sum_{k=p+1}^{\infty} E \left[ \Phi(t, k) \alpha_{S_{t-k}'} \alpha_{S_{t-p+1}}' \Phi(t, p)|S^t_{t-p+1} \right] = \prod_{i=1}^{p} \Phi_{S_{t-i+1}'} J_{S_{t-p+1}} \sum_{k=p+1}^{\infty} B_{1k-p+1} J'_{S_{t-p+1}} (\prod_{i=1}^{p} \Phi_{S_{t-i+1}})'$$

28
To solve for $B_1 = \sum_{k=p+1}^{\infty} B_{1k-p+1}$, we observe that:

\[ J_1 B_1 J' = J_1 (M \otimes I_n) \Phi B_0 \alpha' J' \]

\[ J_K B_1 J'_K = J_K (M \otimes I_n) \Phi B_0 \alpha' J' \]

which implies $B_1 J' = (M \otimes I_n) \Phi B_0 \alpha' J'$.

Next we consider the case $j = p+1$ focusing on $\sum_{k=p+2}^{\infty} E \left[ \Phi(t, k) \alpha_{S_{t-k}} \alpha'_{S_{t-p+1}} \Phi(t, p+1) | S_{t-p+1}^t \right]$.

For $j = p+1$ and $k = p+2$:

\[ E \left[ \Phi(t, k) \alpha_{S_{t-k}} \alpha'_{S_{t-p+1}} \Phi(t, p+1) | S_{t-p+1}^t \right] = \]

\[ = \prod_{i=1}^{p} \Phi_{S_{t-i+1}} J_{S_{t-p+1}} (M \otimes I_n) \Phi B_{12} \Phi' J' (\prod_{i=1}^{p} \Phi_{S_{t-i+1}})' \]

\[ = \prod_{i=1}^{p} \Phi_{S_{t-i+1}} J_{S_{t-p+1}} B_{23} J'_{S_{t-p+1}} (\prod_{i=1}^{p} \Phi_{S_{t-i+1}})' \]

where the $nK \times nK$ matrix $B_{23}$ is defined as $B_{23} = diag_K \{(M \otimes I_n) \Phi B_{12} \Phi' J' \}$.

For $j = p+1$ the generic term $k > p+2$ is such as:

\[ E \left[ \Phi(t, k) \alpha_{S_{t-k}} \alpha'_{S_{t-p+1}} \Phi(t, p+1) | S_{t-p+1}^t \right] = \]

\[ = \prod_{i=1}^{p} \Phi_{S_{t-i+1}} J_{S_{t-p+1}} (M \otimes I_n) \Phi B_{1k-p} \Phi' J' (\prod_{i=1}^{p} \Phi_{S_{t-i+1}})' \]

\[ = \prod_{i=1}^{p} \Phi_{S_{t-i+1}} J_{S_{t-p+1}} B_{2k-p+1} J'_{S_{t-p+1}} (\prod_{i=1}^{p} \Phi_{S_{t-i+1}})' \]

where the $nK \times nK$ matrix $B_{2k-p+1}$ verifies the recursion $B_{2k-p+1} = diag_K \{(M \otimes I_n) \Phi B_{1k-p} \Phi' J' \}$.

Hence for $j = p+1$ the infinite sum verifies:

\[ \sum_{k=p+2}^{\infty} E \left[ \Phi(t, k) \alpha_{S_{t-k}} \alpha'_{S_{t-p+1}} \Phi(t, p+1) | S_{t-p+1}^t \right] = \prod_{i=1}^{p} \Phi_{S_{t-i+1}} J_{S_{t-p+1}} \sum_{k=p+2}^{\infty} B_{2k-p+1} J'_{S_{t-p+1}} (\prod_{i=1}^{p} \Phi_{S_{t-i+1}})' \]

To solve for $B_2 = \sum_{k=p+2}^{\infty} B_{2k-p+1}$ we consider the system:

\[ J_1 B_2 J' = J_1 (M \otimes I_n) \Phi B_1 \Phi' J' \]

\[ \vdots \]

\[ J_K B_2 J'_K = J_K (M \otimes I_n) \Phi B_1 \Phi' J' \]
which yields:

\[ B_2 J' = (M \otimes I_n) \Phi B_1 \Phi' J' \]

Finally for each value of \( j = p + 2, p + 3, \ldots \), each sum over \( k > j \) verifies:

\[
\sum_{k=j+1}^{\infty} E \left[ \Phi(t, k) \alpha_{S_{t-k}} \alpha'_{S_{t-p+1}} \Phi(t, p + 1) | S_{t-p+1}^t \right] = \prod_{i=1}^{p} \Phi_{S_{t-i+1}} B_{j-p+1} J_{S_{t-p+1}} J_{S_{t-p+1}} (\prod_{i=1}^{p} \Phi_{S_{t-i+1}})' \]

where \( B_{j-p+1} = \sum_{k=j+1}^{\infty} B_{j-p+1} k \) and

\[
B_{j-p+1} J' = (M \otimes I_n) \Phi B_{j-p} \Phi' J' \]

Defining \( B = \sum_{j=1}^{\infty} B_j \), we have:

\[
BJ' = (M \otimes I_n) \Phi B_0 \alpha' J' + (M \otimes I_n) \Phi B \Phi' J' \]

whose solution is given in (3.10). This yields (3.7) and thus completes the proof. \( \blacksquare \)

**Proof of Theorem 2** (a) Using the law of iterated expectations, the given regime third central moment verifies:

\[
E \left[ \prod_{t=\ell_{i,j,k}} (x_{\ell t} - E(x_{\ell t} | S_{t-p+1}^t)) \left| S_{t-p+1}^t \right| \right] = \]

\[
= \sum_{S_{t-q}^t} P(S_{t-q}^t | S_{t-p+1}^t) E \left[ \prod_{t=\ell_{i,j,k}} (x_{\ell t} - E(x_{\ell t} | S_{t-q}^t) + E(x_{\ell t} | S_{t-q}^t) - E(x_{\ell t} | S_{t-p+1}^t)) | S_{t-q}^t \right] \]

\[
= \sum_{S_{t-q}^t} P(S_{t-q}^t | S_{t-p+1}^t) E \left[ \prod_{t=\ell_{i,j,k}} (x_{\ell t} - E(x_{\ell t} | S_{t-q}^t)) \right] + \]

\[
+ \sum_{(\ell_1, \ell_2, \ell_3) \in \mathcal{P}_1} \prod_{t=\ell_1, \ell_2} (x_{\ell t} - E(x_{\ell t} | S_{t-q}^t)) (E(x_{\ell t} | S_{t-q}^t) - E(x_{\ell t} | S_{t-p+1}^t)) \]

\[
+ \sum_{(\ell_1, \ell_2, \ell_3) \in \mathcal{P}_1} \prod_{t=\ell_1, \ell_2} (E(x_{\ell t} | S_{t-p+1}^t) - E(x_{\ell t} | S_{t-q}^t)) (x_{\ell t} - E(x_{\ell t} | S_{t-q}^t)) \]

\[
+ \prod_{t=\ell_{i,j,k}} (E(x_{\ell t} | S_{t-q}^t) - E(x_{\ell t} | S_{t-p+1}^t)) | S_{t-q}^t \]

where \( \mathcal{P}_1 = \{(i, j, k), (i, k, j), (j, k, i)\} \). The first term in the sum above represents \( \eta_{ijk}(S_{t-q}^t) \) whereas the third term is null. Expression (4.1) is straightforwardly obtained. \( \blacksquare \)
Proof of Theorem 2 (b) The fourth central moment conditional on regime verifies:

\[ E \left[ \prod_{\ell=i,j,k,l} (x_{\ell t} - E(x_{\ell t} | S_{t-p}^t)) \right] = \]

\[ = \sum_{S_{t-q}^t} P(S_{t-q}^t | S_{t-p}^t) E \left[ \prod_{\ell=i,j,k,l} (x_{\ell t} - E(x_{\ell t} | S_{t-q}^t)) + E(x_{\ell t} | S_{t-q}^t) - E(x_{\ell t} | S_{t-p}^t) \right] S_{t-q}^t \]

\[ = \sum_{S_{t-q}^t} P(S_{t-q}^t | S_{t-p}^t) E \left[ \prod_{\ell=i,j,k,l} (x_{\ell t} - E(x_{\ell t} | S_{t-q}^t)) \right] + \]

\[ + \sum_{(\ell_1, \ell_2, \ell_3, \ell_4) \in Q} \prod_{\ell = \ell_1, \ell_2} (x_{\ell t} - E(x_{\ell t} | S_{t-q}^t)) \prod_{\ell = \ell_3, \ell_4} (E(x_{\ell t} | S_{t-q}^t) - E(x_{\ell t} | S_{t-p}^t)) \]

\[ + \sum_{(\ell_1, \ell_2, \ell_3, \ell_4) \in P_3} \prod_{\ell = \ell_1, \ell_2} (E(x_{\ell t} | S_{t-p}^t) - E(x_{\ell t} | S_{t-q}^t)) (x_{\ell t} - E(x_{\ell t} | S_{t-q}^t)) \]

\[ + \prod_{\ell = i, j, k, l} (E(x_{\ell t} | S_{t-q}^t) - E(x_{\ell t} | S_{t-p}^t)) S_{t-q}^t \] \hspace{1cm} (A.3)

where \( Q = \{(i, j, k, l), (i, j, l, k), (i, k, l, j), (j, k, l, i)\} \) and \( P_3 = \{(i, j, k, l), (i, k, j, l), (i, l, j, k), (j, k, l, i), (j, l, i, k), (k, l, i, j)\} \). By definition of \( \kappa_{ijkl}(S_{t-q}^t) \) the first term in the right-hand-side above verifies:

\[ E \left[ \prod_{\ell=i,j,k,l} (x_{\ell t} - E(x_{\ell t} | S_{t-q}^t)) \right] S_{t-q}^t \] = \( \kappa_{ijkl}(S_{t-q}^t) \) + \]

\[ + \sum_{(\ell_1, \ell_2, \ell_3, \ell_4) \in P_2} E \left[ \prod_{\ell = \ell_1, \ell_2} (x_{\ell t} - E(x_{\ell t} | S_{t-q}^t)) \right] E \left[ \prod_{\ell = \ell_3, \ell_4} (x_{\ell t} - E(x_{\ell t} | S_{t-q}^t)) \right] S_{t-q}^t \] \hspace{1cm} (A.4)

where \( P_2 = \{(i, j, k, l), (i, k, j, l), (i, l, j, k)\} \). It is easily seen that the fourth term in (A.3) is null. Plugging (A.4) into (A.3) and cancelling the fourth term yields the expected result. \( \blacksquare \)
Table 1 First two annualized moments of UK stocks and bonds excess returns

<table>
<thead>
<tr>
<th></th>
<th>Stocks</th>
<th>Bonds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean returns</td>
<td>-38.49</td>
<td>19.16</td>
</tr>
<tr>
<td>Volatilities</td>
<td>25.30</td>
<td>12.47</td>
</tr>
<tr>
<td>Correlation</td>
<td>-.03</td>
<td>.50</td>
</tr>
<tr>
<td>stocks/bonds</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Model parameters taken from GT (2005); sample dates 1976-2/2000-12; uncond. stands for unconditional; mean returns are annualized via the compound interest formula; volatilities are annualized by multiplying the monthly standard deviation by $\sqrt{12}$. 

32
<table>
<thead>
<tr>
<th></th>
<th>Stocks</th>
<th>Bonds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skewness</td>
<td>.02</td>
<td>.00</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3.02</td>
<td>3.00</td>
</tr>
<tr>
<td>Co-skewness</td>
<td>-.06</td>
<td>.00</td>
</tr>
<tr>
<td>Co-kurtosis</td>
<td>-.02</td>
<td>1.51</td>
</tr>
</tbody>
</table>

Notes: Model parameters taken from GT (2005); sample dates 1976-1/2000-12; Uncond. stands for unconditional. All moments in Table 2 have been calculated with $q = 10$. 

33
Table 3 Higher-order moments in Example 5

<table>
<thead>
<tr>
<th></th>
<th>IP</th>
<th>E</th>
<th>Mix</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S</td>
<td>K</td>
<td>S(E,IP)</td>
</tr>
<tr>
<td>Empirical:</td>
<td>-0.93 6.99</td>
<td>-0.43 5.05</td>
<td>-0.69 3.57</td>
</tr>
<tr>
<td>Model-based:</td>
<td>-0.31 3.24</td>
<td>-0.10 3.05</td>
<td>-0.22 1.62</td>
</tr>
<tr>
<td>Under normality:</td>
<td>0 3.00</td>
<td>0 3.00</td>
<td>0 1.75</td>
</tr>
</tbody>
</table>

Notes: IP refers to the US Industrial Production Index and E to the US non-farm employment; S and K stand for univariate skewness and kurtosis; $S(E,IP)$ denotes the co-skewness between E and the square of IP as in (4.6); $K_{22}(E,IP)$ denotes the co-kurtosis 2-2 defined in (6.4).
Figure 1: Stocks and bonds series plus mean given the most likely regime
Figure 2: Approximation and true value of kurtosis for different values of $q$

Notes: parameter values refer to model (4.5); on the x-axis $q$ ranges from 1 to 15; the continuous line shows the true kurtosis value and the dashed line the approximated value.
Figure 3: Kernel density estimate $\hat{f}(x_t)$ (—), two-component mixture $f_1(x_t)$ (·—), four-component mixture $f_2(x_t)$ (· · ·), eight-component mixture $f_3(x_t)$ (——)
Figure 4: Density estimates

Notes: in each panel the dotted lines refer to the 1% quantile, the mode, and the 99% quantile; all density estimates are obtained with \( q \) set equal to ten.