Communications in Statistics - Theory and Methods

ISSN: 0361-0926 (Print) 1532-415X (Online) Journal homepage: https://www.tandfonline.com/loi/lsta20

Marginal distribution of Markov-switching VAR processes

Gabriele Fiorentini, Christophe Planas & Alessandro Rossi

To cite this article: Gabriele Fiorentini, Christophe Planas & Alessandro Rossi (2017) Marginal distribution of Markov-switching VAR processes, Communications in Statistics - Theory and Methods, 46:13, 6605-6623, DOI: 10.1080/03610926.2015.1132324

To link to this article: https://doi.org/10.1080/03610926.2015.1132324

Published online: 13 Mar 2017.

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
2017, VOL. 46, NO. 13, 6605–6623
http://dx.doi.org/10.1080/03610926.2015.1132324

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS

Marginal distribution of Markov-switching VAR processes

Gabriele Fiorentini[a,b], Christophe Planas[c], and Alessandro Rossi[c]

[a]Rimini Centre for Economic Analysis, Rimini, Italy; [b]University of Florence, Florence, Italy; [c]European Commission, Joint Research Centre, Ispra (VA), Italy

ABSTRACT
We make available simple and accurate closed-form approximations to the marginal distribution of Markov-switching vector auto-regressive (MS VAR) processes. The approximation is built upon the property of MS VAR processes of being Gaussian conditionally on any semi-infinite sequence of the latent state. Truncating the semi-infinite sequence and averaging over all possible sequences of that finite length yields a mixture of normals that converges to the unknown marginal distribution as the sequence length increases. Numerical experiments confirm the viability of the approach which extends to the closely related class of MS state space models. Several applications are discussed.

ARTICLE HISTORY
Received 7 March 2015
Accepted 11 December 2015

KEYWORDS
Markov-switching; mixtures of normals; state space models; VAR models.

MATHEMATICS SUBJECT CLASSIFICATION
C1

1. Introduction

Markov-switching vector auto-regressive (MS VAR) models have proven most useful to detect non linearities due to interventions, structural changes, and abnormal events. Thanks to their ability to fit time series subject to changes in pattern, these models have entertained a large success in empirical works over a wide variety of disciplines. They have been applied for instance in macroeconomics to the analysis of the business cycle (Hamilton, 1989; McConnell and Perez-Quiros, 2000) and monetary policy (Sims and Zha, 2006), in finance to asset pricing (Cecchetti et al., 1990, 1993), in environmental science to characterize wind time series (Ailliot and Monbet, 2012), in medicine for clinical monitoring (Gordon and Smith, 1990), in speech recognition (Juang and Rabiner, 1985), as well as in many other fields. General discussions and additional references can be found in West and Harrison (1997), Kim and Nelson (1999), Scott (2002), Fruhwirth-Schnatter (2006), and Ang and Timmermann (2011).

The statistical properties of MS VAR models have been analyzed by, among others, Timmermann (2000), Yang (2000), Francq and Zakoian (2001, 2002), Stelzer (2009), and Fiorentini, Planas, and Rossi (2015). These studies center on stationarity issues, on the unconditional mean and variance, and on skewness and kurtosis. In this paper, we complete this stream by focusing on the marginal distribution. Knowledge of the marginal distribution of MS VAR processes is advantageous in several circumstances. By summarizing the non linear features of the model, the marginal law sheds light on noteworthy characteristics such as multi-modalities and regions of probability mass concentration. A comparison with the empirical distribution offers a simple yet powerful tool for model validation. The marginal

CONTACT Alessandro Rossi alessandro.rossi1@ec.europa.eu European Commission, Joint Research Centre, Ispra (VA), Italy.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lsta.

distribution also gives all necessary information for long-term forecasting. It can thus help analysts involved in risk management to infer the value-at-risk over long horizons (Cotter, 2007). Since it enters the likelihood function via the distribution of the first observation, knowledge of the marginal law makes possible the exact calculation of the likelihood function, thus enhancing inference. This is particularly relevant in Bayesian analysis via Markov chain Monte Carlo methods where parameters and latent variables need to be sampled.

In spite of its usefulness, the marginal law of MS VAR processes is still unknown. The approximation we offer exploits the property of MS VAR models to be Gaussian conditionally on any semi-infinite sequence of the latent discrete state. Truncating the semi-infinite sequence to a finite length and averaging over all possible sequences of that length yields a mixture of normals that converges to the unknown marginal distribution as the sequence length increases. The mean and covariance matrix of each normal mixture component are set equal to the first two moments of the MS VAR process given the finite sequence of discrete states, while the probability of occurrence of the sequences gives the mixing weights. The approximation applies as well to the closely related class of MS state space (MS SS) models which have been popularized in the statistical literature by Harrison and Stevens (1976). Numerical experiments confirm the viability of the approach.

The paper is structured as follows. The general MS VAR and MS SS framework is presented in Sec. 2 together with assumptions and notations. We focus on models with finite number of states and time-invariant transition probabilities. Approximating mixtures for the marginal distribution of MS processes are derived in Sec. 3. The accuracy of the approximation is analyzed through a Monte Carlo experiment. Section 4 reviews applications where knowledge of the marginal law enhances inference. Section 5 concludes the paper. All proofs are gathered in the Appendix. A Matlab code that implements the results shown in the paper is available from the authors.

2. Model and assumptions

Let $\{\varepsilon_t\}$ be n-dimensional Gaussian white noise and $\{S_t\}$ an homogeneous K-state irreducible Markov chain defined at discrete time t. The first-order MS VAR process is generated by the stochastic difference equation:

$$x_t = \alpha_{S_t} + \Phi_{S_t}\, x_{t-1} + \Lambda_{S_t}\, \varepsilon_t \tag{2.1}$$

where $x_t = (x_{1t}, \ldots, x_{nt})'$. The $n \times 1$ vector α_{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. 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 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).

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

$$y_t = a_{S_t} + H_{S_t}\, x_t + \gamma_{S_t}\, u_t \tag{2.2}$$

Equations (2.1) and (2.2) make up an MS SS model. We assume that the $m \times 1$ vector u_t is a Gaussian standard white noise independent of ϵ_t. Like in (2.1), the $m \times 1$ vector a_{S_t}, the $m \times n$ matrix H_{S_t}, and the $m \times m$ matrix γ_{S_t} take K different values depending on the realization of the discrete latent variable S_t.

Further notations are needed. For any generic variable z_t, we denote $z_s^t = (z_s, z_{s+1}, \dots, z_t)$ and $z^t = (z_1, z_2, \dots, z_t)$. We also denote by 1_K the $K \times 1$ vector with all elements equal to 1, e_ℓ the $K \times 1$ unit vector $e_\ell = [0'_{\ell-1}, 1, 0'_{K-\ell}]'$ for $\ell = 1, \dots, K$, 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 $\pi_k = P(S_t = k)$ represents the probability of being in state k. For $k = 1, \dots, K$, 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 $n_1K \times n_2$ matrix Q as

$$diag_K Q = diag_K \begin{bmatrix} Q_1 \\ \vdots \\ Q_K \end{bmatrix} = \begin{bmatrix} Q_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & Q_K \end{bmatrix}_{n_1K\times n_2K}$$

where the blocks Q_k are $n_1 \times n_2$ matrices. The $nK \times K$ matrix α and the $nK \times nK$ matrices Λ and Φ are defined accordingly:

$$\alpha = diag_K \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_K \end{bmatrix} \quad \Lambda = diag_K \begin{bmatrix} \Lambda_1 \\ \vdots \\ \Lambda_K \end{bmatrix} \quad \text{and} \quad \Phi = diag_K \begin{bmatrix} \Phi_1 \\ \vdots \\ \Phi_K \end{bmatrix}$$

Note that the matrix J' inverts the $diag_K$ operator, i.e., for any $nK \times n$ matrix Q, $diag_K QJ' = Q$. 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 $vec(A) = Hvec(AJ')$ for the $n^2K^2 \times n^2K$ matrix H such that $H = \sum_{k=1}^K J'_k \otimes (J'_kJ_k)$.*

3. The marginal distribution

There are two special cases where the marginal distribution of the MS VAR process (2.1) is exactly known: (i) there is no autoregressive term, i.e., $\Phi_{S_t} = 0$, and (ii) the model is parameterized in terms of conditional mean instead of intercept, and both the autoregressive coefficients and the shock-loading matrices are time-invariant, i.e., $x_t - \mu_{S_t} = \Phi\,(x_{t-1} - \mu_{S_{t-1}}) + \Lambda\epsilon_t$. These two cases are important as (i) arises in many empirical finance applications due to the efficient market hypothesis (see Ang and Timmermann, 2011) and (ii) corresponds to the model proposed by Hamilton (1989). In these two cases the marginal distribution of x_t can be written as

$$f(x_t) = \sum_{S_t} P(S_t)\phi(x_t; E(x_t|S_t), V(x_t|S_t)) \tag{3.1}$$

where $\phi(\cdot; E, V)$ denotes the normal density with mean E and variance-covariance matrix V. The conditional moments involved in (3.1) are such that $E(x_t|S_t) = \alpha_{S_t}$ and $V(x_t|S_t) = \Lambda_{S_t}\Lambda'_{S_t}$ in the case (i), while in the case (ii) $E(x_t|S_t) = \mu_{S_t}$ and $vec[V(x_t|S_t)] = [I_{n^2} - \Phi \otimes \Phi]^{-1}\Lambda\Lambda'$ (see also Albert and Chib, 1993).

In the unrestricted general case the marginal distribution of the MS VAR process is unknown. However, conditionally on semi-infinite realizations of the discrete latent variable, the MS VAR process is Gaussian so the marginal distribution verifies the limit expression:

$$f(x_t) = \lim_{p\to\infty} \sum_{S^t_{t-p+1}} P\left(S^t_{t-p+1}\right) \phi\left(x_t; E\left(x_t|S^t_{t-p+1}\right), V\left(x_t|S^t_{t-p+1}\right)\right) \tag{3.2}$$

As integrating over all possible semi-infinite regime paths is numerically infeasible, no direct use can be made of the expression (3.2). Rather, for any integer $p > 0$, we define $f_p(\cdot)$ as the finite mixture of K^p-normals:

$$f_p(x_t) = \sum_{S^t_{t-p+1}} P\left(S^t_{t-p+1}\right) \phi\left(x_t; E\left(x_t|S^t_{t-p+1}\right), V\left(x_t|S^t_{t-p+1}\right)\right) \tag{3.3}$$

The Gaussian mixture $f_p(x_t)$ has several properties: (a) it conserves the first two unconditional moments of x_t for all $p > 0$, i.e., if $E_p(\cdot)$ denotes the unconditional moment under $f_p(x_t)$, then $E_p(x_t) = E(x_t)$ and $E_p(x_t x_t') = E(x_t x_t')$; (b) it reproduces exactly the marginal distribution $f(x_t)$ when the conditional density $f(x_t|S^t_{t-p+1})$ is Gaussian, i.e., $f(x_t|S^t_{t-p+1}) = \phi(x_t; E(x_t|S^t_{t-p+1}), V(x_t|S^t_{t-p+1}))\forall S^t_{t-p+1} \Rightarrow f_p(x_t) = f(x_t)$; and (c) it converges to the marginal distribution $f(x_t)$ as p increases, i.e., $\lim_{p\to\infty} f_p(x_t) = f(x_t)$. The mixture $f_p(x_t)$ can thus be expected to provide an accurate approximation to the unknown marginal law $f(x_t)$.

To evaluate $f_p(x_t)$, the conditional moments $E(x_t|S^t_{t-p+1})$ and $V(x_t|S^t_{t-p+1})$ are necessary. Results for the simplest case $p = 1$ can be found in Francq and Zakoian (2001); see also Lemma 2 below. Given $E(x_t|S_t)$ and $V(x_t|S_t)$, the conditional moments $E(x_t|S^t_{t-p+1})$ and $V(x_t|S^t_{t-p+1})$ can be obtained by iterating for $q = 2, 3, \ldots$ until $q = p$ the recursions:

$$E(x_t \mid S_t = \imath_1, S_{t-1} = \imath_2, \ldots, S_{t-q+1} = \imath_q) = \alpha_{\imath_1} + \Phi_{\imath_1} E(x_t \mid S_t = \imath_2, \ldots, S_{t-q+2} = \imath_q)$$

and

$$\begin{aligned} &E(x_t x_t' \mid S_t = \imath_1, S_{t-1} = \imath_2, \ldots, S_{t-q+1} = \imath_q) \\ &\quad = \alpha_{\imath_1}\alpha_{\imath_1}' + \Lambda_{\imath_1}\Lambda_{\imath_1}' + \Phi_{\imath_1} E(x_t x_t' \mid S_t = \imath_2, \ldots, S_{t-q+2} = \imath_q)\Phi_{\imath_1}' \\ &\qquad + \alpha_{\imath_1} E(x_t' \mid S_t = \imath_2, \ldots, S_{t-q+2} = \imath_q)\Phi_{\imath_1}' + \Phi_{\imath_1} E(x_t \mid S_t = \imath_2, \ldots, S_{t-q+2} = \imath_q)\alpha_{\imath_1}' \end{aligned} \tag{3.4}$$

where the realizations $\imath_j$, $j = 1, 2, \ldots, q$ take values $1, 2, \ldots, K$.

These recursions are appealing when the conditional moments are needed for all sequences of increasing length, i.e., $S_t, S^t_{t-1}, \ldots, S^t_{t-p+1}$. Otherwise, it is preferable to use the general closed-form expressions that are given in the following lemma.

Lemma 2. *The first and second moment of x_t conditionally on the sequence S^t_{t-p+1} of length p, $p > 0$, verify:*

(a)

$$\begin{aligned} &E\left(x_t \mid S^t_{t-p+1}\right) \\ &\quad = \alpha_{S_t} + \sum_{i=2}^{p}\prod_{j=2}^{i} \Phi_{S_{t-j+2}}\alpha_{S_{t-i+1}} + \prod_{j=1}^{p} \Phi_{S_{t-j+1}} J_{S_{t-p+1}} [I_{nK} - (M\otimes I_n)\Phi]^{-1}(M\otimes I_n)\alpha 1_K \end{aligned} \tag{3.5}$$

(b)

$$E\left(x_t x_t' \mid S_{t-p+1}^t\right) = \alpha_{S_t}\alpha_{S_t}' + \Lambda_{S_t}\Lambda_{S_t}' + \sum_{j=1}^{p-1}\prod_{i=1}^{j}\Phi_{S_{t-i+1}}(\alpha_{S_{t-j}}\alpha_{S_{t-j}}' + \Lambda_{S_{t-j}}\Lambda_{S_{t-j}}') \times \left(\prod_{i=1}^{j}\Phi_{S_{t-i+1}}\right)' + \prod_{i=1}^{p}\Phi_{S_{t-i+1}}J_{S_{t-p+1}}AJ_{S_{t-p+1}}'\left(\prod_{i=1}^{p}\Phi_{S_{t-i+1}}\right)' + \Gamma + \Gamma' \tag{3.6}$$

where the $nK \times nK$ block-diagonal matrix A is such as

$$vec(A) = H\{I_{n^2K} - [J\Phi \otimes ((M \otimes I_n)\Phi)]H\}^{-1}vec(A_0J'), \tag{3.7}$$

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:

$$A_0 = diag_K\{(M \otimes I_n)(\alpha\alpha' + \Lambda\Lambda')J'\} \tag{3.8}$$

The $n \times n$ matrix Γ is equal to

$$\Gamma = \sum_{k=1}^{p-1}\prod_{i=1}^{k}\Phi_{S_{t-i+1}}\alpha_{S_{t-k}}\alpha_{S_t}' + \sum_{j=1}^{p-2}\sum_{k=j+1}^{p-1}\prod_{i=1}^{k}\Phi_{S_{t-i+1}}\alpha_{S_{t-k}}\alpha_{S_{t-j}}'\left(\prod_{i=1}^{k}\Phi_{S_{t-i+1}}\right)' \tag{3.9}$$

$$+\prod_{i=1}^{p}\Phi_{S_{t-i+1}}J_{S_{t-p+1}}\left[I_{nK} - (M \otimes I_n)\Phi\right]^{-1}(M \otimes I_n)\alpha 1_K\sum_{j=0}^{p-2}\alpha_{S_{t-j}}'\left(\prod_{i=1}^{j}\Phi_{S_{t-i+1}}\right)' \tag{3.10}$$

$$+\prod_{i=1}^{p}\Phi_{S_{t-i+1}}J_{S_{t-p+1}}B_0 e_{S_{t-p+1}}'\alpha_{S_{t-p+1}}'\left(\prod_{i=1}^{p-1}\Phi_{S_{t-i+1}}\right)' \tag{3.11}$$

$$+\prod_{i=1}^{p}\Phi_{S_{t-i+1}}J_{S_{t-p+1}}BJ_{S_{t-p+1}}'\left(\prod_{i=1}^{p}\Phi_{S_{t-i+1}}\right)' \tag{3.12}$$

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

$$B_0 = diag_K\{[I_{nK} - (M \otimes I_n)\Phi]^{-1}B_{01}1_K\} \tag{3.13}$$

$$\text{with} \qquad B_{01} = diag_K\{(M \otimes I_n)\alpha 1_K\} \tag{3.14}$$

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

$$vec(B) = H\{I_{n^2K} - [J\Phi \otimes ((M \otimes I_n)\Phi)]H\}^{-1}vec((M \otimes I_n)\Phi B_0\alpha'J') \tag{3.15}$$

All quantities involved in Lemma 2 are straightforwardly available from the specification (2.1). Compared to the recursive procedure (3.4), the closed-form formulae (3.5)–(3.15) reduce the total storage by a factor equal to $(K^p - K)/(K-1)$, so Lemma 2 has the advantage of storage minimization. Plugging $E(x_t|S_{t-p+1}^t)$ and $V(x_t|S_{t-p+1}^t)$ into (3.3) yields the density components of $f_p(x_t)$.

Table 1. Parameter values for six MS AR models.

Density	α_1, α_2	Φ_1, Φ_2	Λ_1, Λ_2	p_{11}, p_{22}
#1 Weakly skewed	−0.5, 0.5	0.7, 0.8	$1, \sqrt{2}$	0.8, 0.8
#2 Kurtotic	0, 0	0.9, 0.9	1, 0.2	0.9, 0.9
#3 Strongly skewed	2, 0	0.5, 0.5	0.1, 1	0.8, 0.2
#4 Bimodal	1, −1	0.7, 0.7	1, 1	0.9, 0.9
#5 Separated bimodal	3, −3	0.2, 0.2	1, 1	0.8, 0.8
#6 Asymmetric trimodal	1.5, −5.5	0.5, 0.2	$5, \sqrt{0.06}$	0.9, 0.9

Lemma 2 extends readily to MS SS models, with the first two moments being such as

$$E\left(y_t|S_{t-p+1}^t\right) = c_{S_t} + H_{S_t}E\left(x_t|S_{t-p+1}^t\right)$$
$$V\left(y_t|S_{t-p+1}^t\right) = H_{S_t}V\left(x_t|S_{t-p+1}^t\right)H'_{S_t} + \gamma_{S_t}\gamma'_{S_t} \tag{3.16}$$

These moments can be plugged into formula (3.3) to build the mixture $f_p(y_t)$ which approximates the marginal law of the MS SS process $f(y_t)$.

In the examples below we analyze the quality of the approximation to the true marginal law through a Monte Carlo experiment. First we consider six two-state MS univariate processes with parameter values displayed in Table 1. These models have been chosen to yield marginal densities whose main features are frequently encountered in practice (see Marron and Wand, 1992).

Figure 1 shows histograms of $N = 10^7$ realizations, say $x^{(1)}, \ldots, x^{(N)}$, simulated from the ergodic distribution of each process; i.e., starting from the initial condition $x_{t-500} = 0$, we simulate each process until x_t, collect $x^{(1)} = x_t$, and repeat the procedure N times. The first three models generate marginal distributions that are unimodal: the first histogram is weakly

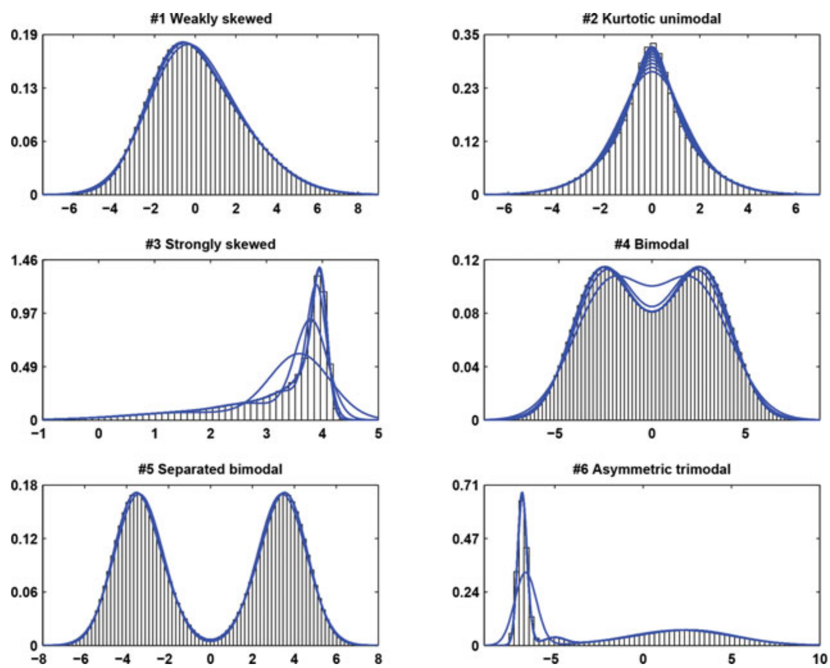

Figure 1. Density approximations. Notes: the histograms are produced from 10^7 simulations of each model; the approximating densities $f_p(x_t)$ are shown by solid lines for $p = 1, \ldots, 10$.

Table 2. Rejection frequencies for Cramer-von Mises test at the 5% level.

p	Model	#1	#2	#3	#4	#5	#6	GT
		Sample size $N = 5,000$						
1		0.397	1.000	1.000	1.000	0.096	1.000	0.068
2		0.091	0.991	1.000	0.204	0.050	0.071	0.053
3		0.053	0.864	0.982	0.060	0.050	0.051	0.046
4		0.050	0.532	0.062	0.053	0.050	0.051	0.045
5		0.050	0.259	0.049	0.052	0.050	0.051	0.045
6		0.050	0.134	0.049	0.052	0.050	0.051	0.045
7		0.050	0.087	0.049	0.052	0.050	0.051	0.044
8		0.050	0.068	0.049	0.052	0.050	0.051	0.044
9		0.050	0.059	0.049	0.052	0.050	0.051	0.044
10		0.050	0.056	0.049	0.052	0.050	0.051	0.044
		Sample size $N = 20,000$						
1		0.985	1.000	1.000	1.000	0.425	1.000	0.192
2		0.252	1.000	1.000	0.927	0.046	0.162	0.077
3		0.073	1.000	1.000	0.090	0.047	0.049	0.056
4		0.052	1.000	0.126	0.051	0.047	0.050	0.053
5		0.047	0.988	0.048	0.050	0.047	0.050	0.052
6		0.047	0.775	0.047	0.050	0.047	0.050	0.051
7		0.047	0.377	0.047	0.050	0.047	0.050	0.051
8		0.047	0.166	0.047	0.050	0.047	0.050	0.050
9		0.047	0.092	0.047	0.050	0.047	0.050	0.050
10		0.047	0.068	0.047	0.050	0.047	0.050	0.050

Notes: # 1 Weakly skewed, # 2 Kurtotic, # 3 Strongly skewed, # 4 Bimodal, # 5 Separate bimodal, # 6 Asymmetric trimodal. The 5% critical value of the CvM test is equal to 0.46119 for the univariate cases 1 to 6 (Csorgo and Faraway, 1996), and to .33088 for the bivariate GT case (Cotterill and Csorgo, 1982); the rejection rates are calculated over 20,000 Monte Carlo simulations.

skewed, the second is heavily kurtotic, and the third one is strongly skewed and sharply peaked. The fourth and fifth histograms are bimodal, the two modes of the fifth one being separated by a region of low probability density. The sixth histogram is trimodal, each mode receiving a different weight. Figure 1 also shows the approximating densities $f_p(x_t)$ for p ranging from one to ten. The approximating densities get closer to the histograms as p increases. In most cases a good matching is obtained with low values of p such as $p = 1, 2$. For the kurtotic case, the convergence to the target distribution is slower but the main features of the histogram seem well captured with $p = 10$.

To statistically assess the quality of the approximation, we simulate 20,000 samples of various lengths N from the ergodic distribution of each process and test the hypothesis that the simulated samples come from the mixture $f_p(x_t)$, $p = 1, 2, \ldots, 10$, using the Cramer–von Mises (CvM) statistics. We record the empirical rejection rate at the usual 5% level of significance; close matching between the empirical rejection rate and the theoretical test size indicates that a sample of length N is not enough to distinguish statistically between the target and the approximating density. In such cases we conclude that $f_p(x_t)$ approximates very well the unknown marginal law. Of course we expect the empirical rejection rate to increase with N and to decrease with p.

Table 2 shows the results for 20,000 simulated samples of length $N = 5,000$ and $N = 20,000$. In five out of the six models, setting p equal to 5 is sufficient to get an approximation which is statistically indistinguishable from the marginal density according to the CvM statistic although the samples contain a number of observations as large as $N = 20,000$. With shorter samples of $N = 5000$ observations, setting $p = 4$ is sufficient to get Gaussian mixtures which are indistinguishable from the target density. The kurtotic unimodal process is the only case where convergence of the empirical power to the actual size, as p increases, is slow: with $N = 20,000$ observations, using $p = 10$ yields a rejection rate of 7%. The test significance level is reached with $p = 10$ when the sample size is 5000. Overall, the mixture density (3.2) with moderate values of p approximates reasonably well the complex and highly non Gaussian marginal distributions considered in the experiment.

To gauge the quality of the approximation in a multivariate setting we use the model estimated by Guidolin and Timmermann (GT, 2005) who fit a MS VAR model to the UK stock and bond monthly excess returns over the period 1976–2 to 2000–12. They consider three regimes that impact the intercept, the autoregressive matrix, and the variance-covariance matrix of the shocks. GT interpret the regimes as bear, normal, and bull market periods with steady state probabilities equal to 13%, 68%, and 19%, respectively. To approximate the ergodic bivariate distribution of stocks and bonds we build the Gaussian mixture (3.3) using the parameter values reported in Table 4 of GT's paper. Like in Example 1 we evaluate the quality of the approximation by implementing the CvM test on simulated observations. We actually resort to the multivariate version of the CvM test discussed in Cotterill and Csorgo (1982). The last column of Table 2 labeled GT shows the empirical rejection frequencies over 20,000 bivariate samples of length $N = 5000$ and $N = 20,000$. With samples of dimension $N = 20,000$, setting p equal to 3 is enough to get an approximating mixture which is indistinguishable from the true distribution according to the CvM statistic. Setting p equal to 2 appears to be sufficient for samples that do not exceed 5,000 points.

4. Applications

In this section, we discuss several circumstances where knowledge of the marginal distribution of MS VAR and MS SS processes enhances inference.

Likelihood evaluation and filtering in MS VAR models

Exact evaluation of the likelihood function of MS VAR models requires knowledge of the marginal distribution. This appears clearly when factorizing the joint distribution of a sample $(x_1, x_2, \ldots, x_T)$ into the contribution of the innovations times the density weight attached to the first observation:

$$f(x_1, x_2, \ldots, x_T) = f(x_1)\prod_{t=2}^{T} f(x_t|x^{t-1})$$

where the conditioning on model parameters is omitted. Ignorance of $f(x_1)$ leads to consider the conditional likelihood $f(x_2, \ldots, x_T|x_1)$. In regression model with autoregressive errors, Beach and MacKinnon (1978) argue that, albeit disregarding the contribution of the first observation makes no difference asymptotically, in small samples maximizing the exact likelihood yields efficiency gains and small-sample bias reduction. Diebold and Schuermann (1996) reach a similar conclusion in the context of ARCH models. The results in Sec. 3 serve not only to approximate $f(x_1)$ but also to launch the recursions that are needed to evaluate the density of the innovation in the second time-period since:

$$\begin{aligned} f(x_2|x_1) &= \sum_{S_2} f(x_2|S_2, x_1)P(S_2|x_1) \\ &= \sum_{S_2} f(x_2|S_2, x_1)\sum_{S_1} P(S_1|x_1)P(S_2|S_1) \end{aligned}$$

where $P(S_1|x_1) \propto f(x_1|S_1)P(S_1)$. The filtered probability $P(S_1|x_1)$ can be approximated using:

$$f_p(x_1|S_1) = \sum_{S^0_{1-p}} P\left(S^0_{1-p}|S_1\right)\phi\left(x_1; E\left(x_1|S^1_{1-p}\right), V\left(x_1|S^1_{1-p}\right)\right)$$

All necessary ingredients to evaluate the expression above are given in Sec. 3.

Frequentist and Bayesian inference in MS SS models

Inference in MS SS models involves further difficulties that depend on which approach, frequentist or Bayesian, is adopted. Kim (1994) devised the gold standard algorithm for the frequentist approach. To calculate the likelihood function, Kim resorts to pre-sample conditions and additional continuous state variables to be estimated. The factor $f(y_1)$ can instead be approximated as in (3.3) with y_t substituted to x_t and using the moments given in (3.16). Next, evaluating the contribution of the first available innovation $f(y_2|y_1)$ requires initializing the filter for both the discrete and the continuous latent variable, i.e., S_t and x_t. The conditional probability $P(S_0, S_1|y_1) \propto f(y_1|S_0, S_1)P(S_0, S_1)$ can be approximated using

$$f_p(y_1|S_0, S_1) = \sum_{S_{-1},\ldots,S_{-p}} P(S_{-1}, \ldots, S_{-p}|S_0, S_1)\phi\left(y_1; E\left(y_1|S^1_{-p}\right), V\left(y_1|S^1_{-p}\right)\right)$$

The filter for the continuous latent variable can be initialized using $E(x_1|S_1, S_0)$ and $V(x_1|S_1, S_0)$ which are given in Lemma 2. The contribution of the first available innovation then amounts to

$$f(y_2|y_1) = \sum_{S_1}\sum_{S_2} P(S_1, S_2|y_1)f(y_2|S_1, S_2, y_1)$$

where $P(S_1, S_2|y_1) = P(S_2|S_1)P(S_1|y_1)$ is easily derived from $P(S_0, S_1|y_1)$. In Kim's algorithm the term $f(y_2|S_1, S_2, y_1)$ is evaluated as

$$f(y_2|S_1, S_2, y_1) = \phi(y_2; H_{S_2}E(x_2|S_1, S_2, y_1), H_{S_2}V(x_2|S_1, S_2, y_1)H'_{S_2} + \gamma_{S_2}\gamma'_{S_2})$$

The conditional moments $E(x_2|S_1, S_2, y_1)$ and $V(x_2|S_1, S_2, y_1)$ can be obtained by Kalman recursions starting from $E(x_1|S_1, S_0)$ and $V(x_1|S_1, S_0)$ given in Lemma 2.

In the Bayesian context, the discrete latent variable is most efficiently sampled using the algorithm proposed by Gerlach, Carter, and Kohn (GCK, 2000). The GCK algorithm follows a Gibbs scheme to draw the discrete latent variable one-at-a-time from the full conditional distributions:

$$P\left(S_t|y^T, S_1^{t-1}, S_{t+1}^T\right) \propto f\left(y_{t+1}^T|y^t, S^T\right) f\left(y_t|y^{t-1}, S^t\right) P\left(S_t|S_1^{t-1}, S_{t+1}^T\right), \quad t = 1, 2, \ldots, T$$

For $t = 1$, the full conditional $P(S_1|y^T, S_2^T)$ involves $f(y_1|S_1)$ which is generally unknown. Like for the Kim's filter, it can be approximated by

$$f_p(y_1|S_1) = \sum_{S_0,\ldots,S_{1-p}} P\left(S_0, \ldots, S_{1-p}|S_1\right)\phi\left(y_1; E\left(y_1|S^1_{1-p}\right), V\left(y_1|S^1_{1-p}\right)\right)$$

The block-sampler version of the GCK algorithm derived by Fiorentini et al. (2014) benefits similarly from the knowledge of the marginal distribution.

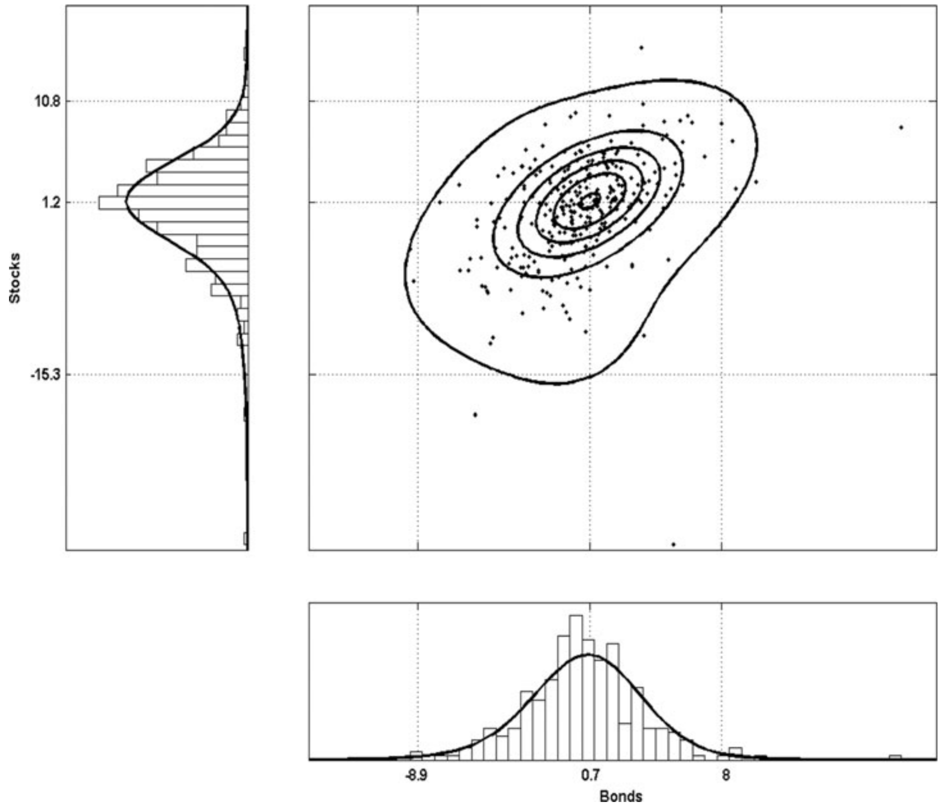

Figure 2. Density estimates for Guidolin and Timmermann (2005) case. Notes: empirical marginal distributions are shown as histograms; the continuous lines show the marginal distribution and the contours of the bivariate distribution obtained by setting $p = 3$ in (3.3) and using the parameter estimates given in Table 4 of GT (2005); the dotted lines refer to the 1% quantile, the mode, and the 99% quantile; the points in the central plot represent the data.

Model validation

The marginal distribution also offers a simple diagnostic tool. Indeed comparing the unconditional distribution implied by the model to the empirical distribution yields insights about the model fit. Let us consider for instance GT's application discussed in Sec. 3. Figure 2 shows the model-based marginal distribution of UK stocks and bonds. The figure shows the contours of the joint distribution as well as the two univariate distributions. The joint distribution appears to be close to normal in the neighbourhood of the mode, the departures being mostly concentrated in the tails. To visualise the model fit, Figure 2 also displays the data points in the central panel as well as histograms of the data in the lateral panels. Several outlying observations are noticeable. The model-based distribution seems broadly congruent with the data: no mismatch indicating mispecification appears.

5. Conclusion

We build mixtures of Gaussians that converge to the true marginal distribution of MS VAR processes as the number of components increases. These mixtures make use of model-based quantities such as the first two moments given a finite sequence of regimes as well as the probability attached to each regime sequence. These results readily extend to the closely related class of MS SS models. The marginal distribution so-obtained has several utilities: it can be used to exactly initialize the filter for the discrete latent variable, to compute the exact likelihood function, and to sample efficiently the discrete latent variable in Bayesian analysis. It also provides a simple yet powerful tool for model validation.

Acknowledgments

We are grateful to Dante Amengual, Massimo Guidolin, Enrique Sentana, an anonymous referee, as well as audiences at CEMFI Madrid, JRC Ispra, RCEF 2014 Rimini, ESEM 2014 Toulouse, and ICEEE 2015 Salerno for their comments and suggestions. Of course, the usual caveat applies. Fiorentini acknowledges funding from MIUR PRIN MISURA – Multivariate models for risk assessment.

References

Ailliot, P., Monbet, V. (2012). Markov-switching autoregressive models for wind time series. *Environ. Modell. Software* 30:92–101.
Albert, J., Chib, S. (1993). Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *J. Bus. Econ. Stat.* 11:1–15.
Ang, A., Timmermann, A. (2011). Regime changes and financial markets. National Bureau of Economic Research, Working Paper 17182.
Beach, C., MacKinnon, J. (1978). A maximum likelihood procedure for regression with autocorrelated errors. *Econometrica* 46(1):51–58.
Bougerol, P., Picard, N. (1992). Strict stationarity of generalized autoregressive processes. *Ann. Probab.* 20:1714–1730.
Brandt, A. (1986). The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Adv. Appl. Probab.* 18:211–220.
Cecchetti, S., Lam, P-S., Mark, N. (1990). Mean reversion in equilibrium asset prices. *Am. Econ. Rev.* 80:398-441.
Cecchetti, S., Lam, P-S., Mark, N. (1993). The equity premium and the risk-free rate. *J. Monetary Econ.* 31:21–45.
Cotter, J. (2007). Varying the VaR for unconditional and conditional environments. *J. Int. Money Finance* 26:1338–1354.
Cotterill, D., Csorgo, M. (1982). On the limiting distribution of and critical values for the multivariate Cramer-von Mises statistic. *Ann. Stat.* 10(1):233–244.
Csorgo, S., Faraway, J. (1996). The exact and asymptotic distributions of Cramer-von Mises statistics. *J. Royal Stat. Soc. B* 58(1):221–234.
Diebold, F., Schuermann, T. (1996). Exact maximum likelihood estimation of observation-driven econometric models. National Bureau of Economic Research, Working Paper 194.
Fiorentini, G., Planas, C., Rossi, A. (2014). Efficient MCMC sampling in dynamic mixture models. *Stat. Comput.* 1:77–90.
Fiorentini, G., Planas, C., Rossi, A. (2015). Skewness and kurtosis of multivariate Markov-switching processes. *Comput. Stat. Data Anal.* forthcoming.
Francq, C., Zakoian, J.M. (2001). Stationarity of multivariate Markov-switching ARMA models. *J. Economet.* 102:339–364.
Francq, C., Zakoian, J.M. (2002). Comments on the paper by Minxian Yang: some properties of vector autoregressive processes with Markov-switching coefficients. *Economet. Theory* 18:815–818.
Fruhwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. New York: Springer.
Gerlach, R., Carter, C., Kohn, R. (2000). Efficient Bayesian inference for dynamic mixture models. *J. Am. Stat. Assoc.* 95:819–828.
Gordon, K., Smith, A. (1990). Modeling and monitoring biomedical time series. *J. Am. Stat. Assoc.* 85:328–337.
Guidolin, M., Timmermann, A. (2005). Economic implications of bull and bear regimes in UK stock and bond returns. *Econ. J.* 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.
Harrison, P., Stevens, C. (1976). Bayesian forecasting. *J. Royal Stat. Soc. Ser. B.* 38:205–247.
Juang, B., Rabiner, L. (1985). Mixture autoregressive hidden Markov models for speech signals. *IEEE Trans. Acoust. Speech Sig. Process.* 33:1404–1413.
Kim, C.-J. (1994). Dynamic linear models with Markov-switching. *J. Economet.* 60:1–22.
Kim, C.-J., Nelson, C.R. (1999). *State-space Models with Regime Switching: Classical and Gibbs Sampling Approaches with Applications*. Cambridge: Massachusetts Institute of Technology Press.
Marron, J., Wand, M. (1992). Exact mean integrated squared error. *Ann. Stat.* 20(2):712–736.
McConnell, M.M., Perez-Quiros, G. (2000). Output fluctuations in the United States: what has changed since the early 1980's. *Am. Econ. Rev.* 90(5):1464–1476.
Scott, S. (2002). Bayesian methods for hidden Markov models: recursive computing in the 21st century. *J. Am. Stat. Assoc.* 97:337–351.
Sims, C.A., Zha, T. (2006). Were there regime switches in U.S. monetary policy?*Am. Econ. Rev.* 96(1):54–81.
Stelzer, R. (2009). On Markov-switching ARMA processes - stationarity, existence of moments, and geometric ergocity. *Economet. Theory* 25:43–62.
Timmermann, A. (2000). Moments of Markov switching models. *J. Economet.* 96:75–111.
West, M., Harrison, P. (1997). *Bayesian Forecasting and Dynamic Models*2nd ed. New York: Springer.
Yang, M. (2000). Some properties of vector autoregressive processes with Markov-switching coefficients. *Economet. Theory* 16:23–43.

Appendix

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

Proof of Lemma 2 (a). Conditionally on current and past history of the discrete latent variable, model (2.1) can be expressed as

$$x_t = \sum_{j=0}^{\infty} \Phi(t, j)(\alpha_{S_t-j} + \Lambda_{S_{t-j}} \epsilon_{t-j}) \tag{A.1}$$

where

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

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

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

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

For $j = p$

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

For $j = p + 1$

$$E\left[\Phi(t, p+1)\alpha_{S_{t-p-1}} \middle| S_{t-p+1}^t\right] = \prod_{i=1}^{p} \Phi_{S_{t-i+1}} \sum_{S_{t-p}} \sum_{S_{t-p-1}} \Phi_{S_{t-p}} \alpha_{S_{t-p-1}}$$

$$\times P(S_{t-p-1}|S_{t-p})P(S_{t-p}|S_{t-p+1})$$
$$= \prod_{i=1}^{p} \Phi_{S_{t-i+1}} J_{S_{t-p+1}} (M \otimes I_n)\Phi(M \otimes I_n)\alpha 1_K$$

For $j > p$ the general term is such as

$$E[\Phi(t,\, j)\alpha_{S_{t-j}} \big| S_{t-p+1}^{t}] = \prod_{i=1}^{p} \Phi_{S_{t-i+1}} J_{S_{t-p+1}} \left[(M \otimes I_n)\Phi\right]^{j-p} (M \otimes I_n)\alpha 1_K$$

Hence, under stationarity, the infinite sum verifies

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

which proves result (a) in Lemma 2. □

Proof of Lemma 2 (b). Expression (A.1) also implies

$$\begin{aligned}
&E(x_t x_t' \big| S_{t-p+1}^{t}) \\
&= \sum_{j=0}^{\infty} E\left[\Phi(t,\, j)(\alpha_{S_t-j} + \Lambda_{S_{t-j}}\epsilon_{t-j})(\alpha_{S_t-j} + \Lambda_{S_{t-j}}\epsilon_{t-j})'\Phi(t,\, j)' \big| S_{t-p+1}^{t}\right] \\
&\quad + \sum_{j=0}^{\infty}\sum_{k=j+1}^{\infty} E\left[\Phi(t,\, k)(\alpha_{S_t-k} + \Lambda_{S_{t-k}}\epsilon_{t-k})(\alpha_{S_t-j} + \Lambda_{S_{t-j}}\epsilon_{t-j})'\Phi(t,\, j)' \big| S_{t-p+1}^{t}\right] \\
&\quad + \sum_{j=0}^{\infty}\sum_{k=j+1}^{\infty} E\left[\Phi(t,\, j)(\alpha_{S_t-j} + \Lambda_{S_{t-j}}\epsilon_{t-j})(\alpha_{S_t-k} + \Lambda_{S_{t-k}}\epsilon_{t-k})'\Phi(t,\, k)' \big| S_{t-p+1}^{t}\right]
\end{aligned}$$

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

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

The sum of the first p terms verifies

$$\begin{aligned}
&\sum_{j=0}^{p-1} E\left[\Phi(t,\, j)(\alpha_{S_t-j} + \Lambda_{S_{t-j}}\epsilon_{t-j})(\alpha_{S_t-j} + \Lambda_{S_{t-j}}\epsilon_{t-j})'\Phi(t,\, j)' \big| S_{t-p+1}^{t}\right] \\
&\quad = \alpha_{S_t}\alpha_{S_t}' + \Lambda_{S_t}\Lambda_{S_t}' + \sum_{j=1}^{p-1}\prod_{i=1}^{j} \Phi_{S_{t-i+1}}(\alpha_{S_{t-j}}\alpha_{S_{t-j}}' + \Lambda_{S_{t-j}}\Lambda_{S_{t-j}}')\left(\prod_{i=1}^{j} \Phi_{S_{t-i+1}}\right)'
\end{aligned}$$

which corresponds to the first row of (3.6).

For $j = p$, the $(p+1)$th term of the sum is such as

$$
\begin{aligned}
&E\left[\Phi(t,p)(\alpha_{S_{t-p}} + \Lambda_{S_{t-p}}\epsilon_{t-p})(\alpha_{S_{t-p}} + \Lambda_{S_{t-p}}\epsilon_{t-p})'\Phi(t,p)' \,\middle|\, S_{t-p+1}^{t}\right] \\
&= \prod_{i=1}^{p}\Phi_{S_{t-i+1}}\left[\sum_{S_{t-p}}(\alpha_{S_{t-p}}\alpha'_{S_{t-p}} + \Lambda_{S_{t-p}}\Lambda'_{S_{t-p}})P(S_{t-p}|S_{t-p+1})\right]\left(\prod_{i=1}^{p}\Phi_{S_{t-i+1}}\right)' \\
&= \prod_{i=1}^{p}\Phi_{S_{t-i+1}}J_{S_{t-p+1}}(M\otimes I_n)(\alpha\alpha' + \Lambda\Lambda')J'\left(\prod_{i=1}^{p}\Phi_{S_{t-i+1}}\right)' \\
&= \prod_{i=1}^{p}\Phi_{S_{t-i+1}}J_{S_{t-p+1}}A_0J'_{S_{t-p+1}}\left(\prod_{i=1}^{p}\Phi_{S_{t-i+1}}\right)'
\end{aligned}
$$

The last equality results from the definition of A_0 given in (3.8).
For $j = p+1$:

$$
\begin{aligned}
&E\left[\Phi(t,p+1)(\alpha_{S_t-p-1} + \Lambda_{S_{t-p-1}}\epsilon_{t-p-1})(\alpha_{S_t-p-1} + \Lambda_{S_{t-p-1}}\epsilon_{t-p-1})'\Phi(t,p+1)'\middle|S_{t-p+1}^{t}\right] \\
&= \prod_{i=1}^{p}\Phi_{S_{t-i+1}}\sum_{S_{t-p}}\Phi_{S_{t-p}}\left[\sum_{S_{t-p-1}}(\alpha_{S_{t-p-1}}\alpha'_{S_{t-p-1}} + \Lambda_{S_{t-p-1}}\Lambda'_{S_{t-p-1}})P(S_{t-p-1}|S_{t-p})\right]\Phi'_{S_{t-p}} \\
&\quad\times P(S_{t-p}|S_{t-p+1})\left(\prod_{i=1}^{p}\Phi_{S_{t-i+1}}\right)' \\
&= \prod_{i=1}^{p}\Phi_{S_{t-i+1}}\sum_{S_{t-p}}\Phi_{S_{t-p}}J_{S_{t-p}}A_0J'_{S_{t-p}}\Phi'_{S_{t-p}}P(S_{t-p}|S_{t-p+1})\left(\prod_{i=1}^{p}\Phi_{S_{t-i+1}}\right)' \\
&= \prod_{i=1}^{p}\Phi_{S_{t-i+1}}J_{S_{t-p-1}}(M\otimes I_n)\Phi A_0\Phi'J'\left(\prod_{i=1}^{p}\Phi_{S_{t-i+1}}\right)'
\end{aligned}
$$

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

$$
\begin{aligned}
&E\left[\Phi(t,p+1)(\alpha_{S_{t-p-1}} + \Lambda_{S_{t-p-1}}\epsilon_{t-p-1})(\alpha_{S_{t-p-1}} + \Lambda_{S_{t-p-1}}\epsilon_{t-p-1})'\Phi(t,p+1)'\middle|S_{t-p+1}^{t}\right] \\
&= \prod_{i=1}^{p}\Phi_{S_{t-i+1}}J_{S_{t-p+1}}A_1J'_{S_{t-p+1}}\left(\prod_{i=1}^{p}\Phi_{S_{t-i+1}}\right)'
\end{aligned}
$$

For $j > p$, the general term verifies

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

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

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

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

$$= \prod_{i=1}^{p}\Phi_{S_{t-i+1}} J_{S_{t-p+1}} \sum_{j=p}^{\infty} A_{j-p} J'_{S_{t-p+1}} \left(\prod_{i=1}^{p}\Phi_{S_{t-i+1}}\right)'$$

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

$$\begin{aligned}
J_{S_{t-p+1}} A_0 J'_{S_{t-p+1}} &= J_{S_{t-p+1}} (M \otimes I_n)(\alpha\alpha' + \Lambda\Lambda')J' \\
J_{S_{t-p+1}} A_1 J'_{S_{t-p+1}} &= J_{S_{t-p+1}} (M \otimes I_n)\Phi A_0 \Phi' J' \\
&\vdots \\
J_{S_{t-p+1}} A_{j-p} J'_{S_{t-p+1}} &= J_{S_{t-p+1}} (M \otimes I_n)\Phi A_{j-p-1} \Phi' J' \\
&\vdots
\end{aligned}$$

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

$$\begin{aligned}
J_1 A J'_1 &= J_1 (M \otimes I_n)(\alpha\alpha' + \Lambda\Lambda')J' + J_1 (M \otimes I_n)\Phi A \Phi' J' \\
&\vdots \\
J_K A J'_K &= J_K (M \otimes I_n)(\alpha\alpha' + \Lambda\Lambda')J' + J_K (M \otimes I_n)\Phi A \Phi' J'
\end{aligned}$$

Using the definition of A_0 in (3.7) the system above can be written as

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

Since $vec(A) = H\text{vec}(AJ')$ where H is defined in Lemma 1, we can easily obtain (3.7).

(ii) To find Γ we split the double-sum in three components:

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

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

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

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

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

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

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

which gives (3.9). The second sum can be solved using (A.2) as in

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

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

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

which gives (3.10).

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

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

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

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

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

where $B_{01}=(M\otimes I_n)\alpha 1_K$ as in (3.14).

For $j=p-1$, $k=p+1$

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

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

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

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

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

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

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

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

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

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

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

$$J_1B_0J_1' = J_1(M\otimes I_n)\alpha 1_K + J_1(M\otimes I_n)\Phi B_0 1_K$$

$$\vdots$$

$$J_KB_0J_K' = J_K(M\otimes I_n)\alpha 1_K + J_K(M\otimes I_n)\Phi B_0 1_K$$

Using the definition of B_{01} in (3.14) the system above can be written as

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

which gives B_0 as in (3.13). This proves the term in (3.11).

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

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

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

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

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

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

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

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

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

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

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

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

$$E\left[\Phi(t, k)\alpha_{S_{t-k}}\alpha'_{S_{t-p+1}}\Phi(t, p)'|S^t_{t-p+1}\right]$$
$$= \prod_{i=1}^{p} \Phi_{S_{t-i+1}} J_{S_{t-p+1}} (M \otimes I_n)\Phi B_{0k-p}\alpha' J' \left(\prod_{i=1}^{p} \Phi_{S_{t-i+1}} \right)'$$
$$= \prod_{i=1}^{p} \Phi_{S_{t-i+1}} J_{S_{t-p+1}} B_{1k-p+1} J'_{S_{t-p+1}} \left(\prod_{i=1}^{p} \Phi_{S_{t-i+1}} \right)'$$

where the $nK \times nK$ matrix B_{1k-p+1} verifies the recursion $B_{1k-p+1} = \text{diag}_K\{(M \otimes I_n)\Phi B_{0k-p}\alpha' J'\}$. Hence, for $j = p$, summing over $k = p + 1, \ldots$ yields

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

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

$$J_1 B_1 J'_1 = J_1(M \otimes I_n)\Phi B_0 \alpha' J'$$
$$\vdots$$
$$J_K B_1 J'_K = J_K(M \otimes I_n)\Phi B_0 \alpha' J'$$

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

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

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

$$E\left[\Phi(t, p+2)\alpha_{S_{t-p-2}}\alpha'_{S_{t-p+1}}\Phi(t, p+1)'\middle|S^t_{t-p+1}\right]$$
$$= \prod_{i=1}^{p} \Phi_{S_{t-i+1}} J_{S_{t-p+1}} (M \otimes I_n)\Phi B_{12}\Phi' J' \left(\prod_{i=1}^{p} \Phi_{S_{t-i+1}} \right)'$$
$$= \prod_{i=1}^{p} \Phi_{S_{t-i+1}} J_{S_{t-p+1}} B_{23} J'_{S_{t-p+1}} \left(\prod_{i=1}^{p} \Phi_{S_{t-i+1}} \right)'$$

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

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

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

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

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:

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

$$= \prod_{i=1}^{p} \Phi_{S_{t-i+1}} J_{S_{t-p+1}} \sum_{k=p+2}^{\infty} B_{2k-p+1} J'_{S_{t-p+1}} \left(\prod_{i=1}^{p} \Phi_{S_{t-i+1}} \right)'$$

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

$$J_1 B_2 J'_1 = J_1 (M \otimes I_n)\Phi B_1 \Phi' J'$$
$$\vdots$$
$$J_K B_2 J'_K = J_K (M \otimes I_n)\Phi B_1 \Phi' J'$$

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

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

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

where $B_{j-p+1} = \sum_{k=j+1}^{\infty} B_{j-p+1\ k-p+1}$ 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 (3.15). This yields (3.12) and thus completes the proof. □