Econometric Theory, **16**, 2000, 23–43. Printed in the United States of America.

SOME PROPERTIES OF VECTOR AUTOREGRESSIVE PROCESSES WITH MARKOV-SWITCHING COEFFICIENTS

MINXIAN YANG
University of New South Wales

Some statistical properties of a vector autoregressive process with Markov-switching coefficients are considered. Sufficient conditions for this nonlinear process to be covariance stationary are given. The second moments of the process are derived under the conditions. The autocovariance matrix decays at exponential rate, permitting the application of the law of large numbers. Under the stationarity conditions, although sharing the "mean-reverting" property with conventional linear stationary processes, the process offers richer short-run dynamics such as conditional heteroskedasticity, asymmetric responses, and occasional nonstationary behavior.

1. INTRODUCTION

Markov-switching (MS) models have recently attracted the attention of several authors. For example, MS models for the U.S. real gross national product (GNP) have been constructed by Hamilton (1989) and Lam (1990) and shown to be able to reveal certain features of the U.S. business cycle that conventional models failed to account for. Engel and Hamilton (1990) have shown that some quarterly exchange rates are better fitted and forecast by an MS model than by a random walk model. Evans and Lewis (1995) have reexamined the long-run Fisher relation by characterizing inflation as an MS process. Further, estimation methods for the MS models are specifically considered by Hamilton (1990) and Kim (1994), and the issues of hypothesis testing in the MS models are discussed by Hansen (1992) and Hamilton (1996).

Given the achievements of the MS models in fitting empirical data, it appears necessary to further understand the theoretical properties of these models. When the MS models are considered as data-generating processes, it is desirable to classify the generated MS processes by the traditional stationary-nonstationary

I thank the co-editor Bruce Hansen and two referees for their constructive comments. I am also grateful to Ron Bewley and Graham Elliott for their helpful suggestions. An early version of this paper was presented at the 1997 Australian Meetings of the Econometric Society at the University of Melbourne. I thank the participants at the presentation for their comments. Financial support from the Australian Research Council is gratefully acknowledged. Address correspondence to: Minxian Yang, School of Economics, University of New South Wales, Sydney 2052, Australia.

© 2000 Cambridge University Press 0266-4666/00 $9.50

https://doi.org/10.1017/S026646660016102X Published online by Cambridge University Press

criterion and to characterize their autocovariance structure. Economic theories often suggest that discrepancies in economic equilibrium relationships should be stationary over time. The known candidates for such stationary discrepancies include stationary ARMA and fractional ARMA processes. Compared with these stationary processes, the MS processes appear more flexible in the sense that structural shifts and short-run asymmetric responses can be described in a systematic way. If stationary, the MS processes will provide a versatile alternative framework to model the discrepancies in economic equilibrium relationships. Therefore, finding and characterizing stationarity conditions for the MS models appear important.

This paper is an attempt to address the issue raised here. A vector autoregressive model with Markov-switching coefficients (VAR-MS) is considered in this paper. The VAR-MS model under consideration is quite general in the sense that it nests many MS models that have been used in the literature. Sufficient conditions for the VAR-MS process to have finite second moments and to be covariance stationary are found. These conditions are given in terms of the mixtures of the Markov-chain transition matrix and VAR coefficients. Roughly, there are two requirements for stationarity: (1) the mean magnitude of the switching coefficient matrix should be strictly less than unity; (2) the Markov chain that governs the switching mechanism should have a stable unconditional distribution. The first requirement can be interpreted as a restriction on the characteristic roots of each VAR in different regimes, and the second is a restriction on the transition matrix of the Markov chain.

Though deriving the moments of nonlinear processes is difficult in general, the structure of the VAR-MS process allows us to acquire its first two moments. It is found that when the process is covariance stationary the decay rate of the autocovariance matrix is exponential, similar to that of a stationary VAR process. In this sense, the VAR-MS process is a *short*-memory process, as opposed to the *long*-memory process introduced by Granger (1980) and Hosking (1981).

An interesting feature of the VAR-MS process is that it can occasionally follow a nonstationary (or explosive) path while maintaining its stationarity in the long run. This could be a desirable feature in modeling time series data such as real interest rates or exchange rates, which are frequently described as $I(1)$ processes but are not necessarily considered to possess a growing variance.

Another issue of interest is the large sample behavior of the MS processes, which is related to statistical inferences on the MS models. Although further inquiries into this issue are needed, this paper's result on the autocovariance structure of the stationary VAR-MS process provides conditions for the process to follow laws of large numbers.

Unlike the conventional VAR model, the "constant term" (or the coefficient matrices) in the VAR-MS model varies according to the Markov chain and serves as an innovation itself. When considering the impulse response of the VAR-MS process to a shock, it is desirable to define a shock as the error of the one-step optimal prediction. A version of Wold's decomposition for the VAR-MS process

https://doi.org/10.1017/S026646660016102X Published online by Cambridge University Press

is given in terms of these shocks. It is also found that the effect of a shock to the process dies out at an exponential rate. Therefore the VAR-MS process is "mean reverting." The one-step optimal prediction error turns out to exhibit certain conditional heteroskedasticity in the sense that its conditional variance is time varying and dependent on the immediate past of the VAR-MS process.

The paper is organized as follows. Section 2 lays out the model, assumptions, and notation. Main results are contained in Section 3. Section 4 presents a simple example to demonstrate the stationarity conditions. Section 5 is a short concluding remark. Section 6 collects all proofs.

2. PRELIMINARIES

Let $\{\Omega, \mathcal{F}, \mathcal{P}\}$ be a probability space on which a first-order K-state Markov chain $\{\mathbf{s}_t\}$ and an n-dimensional vector process $\{\mathbf{u}_t\}$ with integer index t are defined. Let $\mathcal{F}_t$ be the σ-field generated by $\{\mathbf{u}_t, \mathbf{s}_t; \mathbf{u}_{t-1}, \mathbf{s}_{t-1}; \ldots\}$, which summarizes the history of $\{\mathbf{u}_t, \mathbf{s}_t\}$ up to date t.

The transition matrix of $\{\mathbf{s}_t\}$, $P = (p_{ij})'_{K\times K}$, is defined by

$$p_{ij} = \mathcal{P}(\mathbf{s}_t = j \,|\, \mathbf{s}_{t-1} = i, \mathcal{F}_{t-1}) = \mathcal{P}(\mathbf{s}_t = j \,|\, \mathbf{s}_{t-1} = i) \tag{1}$$

for $i, j = 1, \ldots, K$. The Markov chain will also be expressed in the vector form $\xi_t = \mathbf{e}_{\mathbf{s}_t}$, where $\mathbf{e}_j$ is the jth unit vector in R^K. It is well known that the expectation of ξ_t conditional on ξ_{t-1} is given by

$$E(\xi_t \,|\, \xi_{t-1}) = E(\xi_t \,|\, \mathcal{F}_{t-1}) = P\xi_{t-1}. \tag{2}$$

We consider the following n-dimensional process $\{\mathbf{x}_t\}$ that is constructed form $\mathbf{u}_t$ and $\mathbf{s}_t$.

Assumption 1. The process $\{\mathbf{x}_t\}$ is generated by

$$\mathbf{x}_t = \alpha_{\mathbf{s}_t} + \phi_{\mathbf{s}_t}\mathbf{x}_{t-1} + \mathbf{u}_t, \tag{3}$$

where α_j and ϕ_j $(j = 1, \ldots, K)$ are constant matrices of dimensions $n \times 1$ and $n \times n$, respectively, and $\mathbf{u}_t$ satisfies

$$E(\mathbf{u}_t \,|\, \mathbf{s}_t, \mathcal{F}_{t-1}) = 0, \qquad E(\mathbf{u}_t\mathbf{u}'_t \,|\, \mathbf{s}_t, \mathcal{F}_{t-1}) = \lambda_{\mathbf{s}_t}, \tag{4}$$

where λ_j's $(j = 1, \ldots, K)$ are constant semipositive definite matrices of dimension $n \times n$.

Given its history, $\mathbf{x}_t$ is determined by $\mathbf{s}_t$ and $\mathbf{u}_t$ jointly. The dependency of $\mathbf{x}_t$ on the current state $\mathbf{s}_t$ distinguishes it from the conventional autoregressive process. It is nonlinear in the sense that $\mathbf{x}_t$ is a nonlinear function of the innovation processes $\{\mathbf{s}_t\}$ and $\{\mathbf{u}_t\}$. The data-generating process (3) in fact includes all VAR models with MS coefficients, because higher order VAR can always be stacked into the VAR(1) form. It is worth pointing out that the MS models of Hamilton

https://doi.org/10.1017/S026646660016102X Published online by Cambridge University Press

(1989), Lam (1990), Hansen (1992), and Evans and Lewis (1995) are special cases of (3) with $n \geq 2$.

According to (4), $\{\mathbf{u}_t\}$ is a martingale difference with respect to $\{\mathcal{F}_t\}$. In estimation and hypothesis testing, normality is often assumed for MS models. Assumption 1 can be specialized for the normality assumption by writing $\mathbf{u}_t = \lambda_{\mathbf{s}_t}^{1/2}\epsilon_t$ with $\{\epsilon_t\}$ being independent and identically distributed (i.i.d.) $N(0, I_n)$ and independent of $\{\mathbf{s}_t\}$. In this case, $\mathbf{x}_t$ can be expressed as

$$\mathbf{x}_t = (\xi_t' \otimes I_n)(\tilde{\alpha} + \tilde{\phi}\mathbf{x}_{t-1} + \tilde{\lambda}^{1/2}\epsilon_t), \tag{5}$$

where $\tilde{\alpha} = [\alpha_1', \ldots, \alpha_K']'$, $\tilde{\phi} = [\phi_1', \ldots, \phi_K']'$, and $\tilde{\lambda}^{1/2} = [\lambda_1^{1/2}, \ldots, \lambda_K^{1/2}]'$. In view of (5), $\mathbf{x}_t$ is generated by a mixture of K different VAR regimes where the mixing weights are random (determined by the Markov chain ξ_t) and allow exactly one regime to prevail in each period. The conditional probability density function $\text{pdf}(\mathbf{x}_t|\mathbf{s}_t, \mathcal{F}_{t-1})$ is normal (or determined by the pdf of ϵ_t if ϵ_t is not normal). Therefore the conditional pdf of $\mathbf{x}_t$ given the history $\mathcal{F}_{t-1}$

$$\text{pdf}(\mathbf{x}_t|\mathcal{F}_{t-1}) = \sum_{j=1}^{K} \text{pdf}(\mathbf{x}_t|\mathbf{s}_t = j, \mathcal{F}_{t-1})\mathcal{P}(\mathbf{s}_t = j|\mathbf{s}_{t-1}) \tag{6}$$

is a mixture of normals where the weights $\mathcal{P}(\mathbf{s}_t = j|\mathbf{s}_{t-1})$ are the Markov chain's transition probabilities, which are time varying. In what follows, the normal specification $\mathbf{u}_t = \lambda_{\mathbf{s}_t}^{1/2}\epsilon_t$ will not be imposed.

Intuitively, $\mathbf{x}_t$ generated by (3) may be expressed as

$$\mathbf{x}_t = \sum_{j=0}^{\infty} \phi(t,j)(\alpha_{\mathbf{s}_{t-j}} + \mathbf{u}_{t-j}) = \sum_{j=0}^{\infty} \phi(t,j)\mathbf{v}_{t-j}, \tag{7}$$

where $\mathbf{v}_t = \alpha_{\mathbf{s}_t} + \mathbf{u}_t$ and

$$\phi(t,j) = \begin{cases} I_n, & \text{if } j = 0 \\ \phi_{\mathbf{s}_t}\phi_{\mathbf{s}_{t-1}} \ldots \phi_{\mathbf{s}_{t-j+1}}, & \text{if } j > 0 \end{cases} \tag{8}$$

It is easily seen that $\phi(t,j) = \phi(t,i)\phi(t-i, j-i)$ for $i = 0,1,\ldots,j$.

The expressions in (7) and (8) reveal that the impulse response of $\mathbf{x}_t$ to a given innovation $\mathbf{u}_{t-j}$ (not the innovation resulting from $\phi_{\mathbf{s}_{t-j}}$) depends on the path of the Markov chain from $t-j$ to t and therefore is stochastic. This can be a desirable feature in modeling data where certain innovations invoke asymmetric responses in the short run. However, as shown in Theorem 4, the response of stationary $\mathbf{x}_{t+\tau}$ to the "shock" caused by both $\mathbf{u}_t$ and $\mathbf{s}_t$ is actually symmetric in the long run.

Let $\mathcal{L}_2$ stand for the space of all square-integrable random variables defined on $\{\Omega, \mathcal{F}, \mathcal{P}\}$. Let $\|\cdot\|$ denote the 1-norm of a vector (the sum of absolute values of all elements). When applied to matrices, the norm is understood to be the induced matrix norm. Although the choice of norms is not essential to this paper's results (any norm will do), choosing the 1-norm brings out a useful interpretation for stationarity conditions. Let $\mathbf{r}(\cdot)$ denote the radius of a square matrix (the maxi-

https://doi.org/10.1017/S026646660016102X Published online by Cambridge University Press

mum eigenvalue in modulus). Also, let $J = [I_n, \ldots, I_n]_{n \times nK}$, $\mathbf{1} = [1, \ldots, 1]_{1 \times K}$, $\bar{\mathbf{\Lambda}} = \mathbf{\Lambda} + \mathbf{A}\mathbf{A}'$,

$$\mathbf{A} = \begin{bmatrix} \alpha_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \alpha_K \end{bmatrix}_{nK \times K}, \qquad \mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_K \end{bmatrix}_{nK \times nK},$$

and

$$\mathbf{\Phi} = \begin{bmatrix} \phi_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \phi_K \end{bmatrix}_{nK \times nK}, \qquad \hat{\mathbf{\Phi}} = \begin{bmatrix} \|\phi_1\| & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \|\phi_K\| \end{bmatrix}_{K \times K}.$$

The following assumptions will be used to establish the stationarity of $\{\mathbf{x}_t\}$.

Assumption 2.

$$\max\{\mathbf{r}(\hat{\mathbf{\Phi}}^2 P), \mathbf{r}(\hat{\mathbf{\Phi}} P)\} < 1. \tag{9}$$

The preceding inequality requires the radiuses of the mixed coefficient matrices $\hat{\mathbf{\Phi}}^2 P$ and $\hat{\mathbf{\Phi}} P$ to be less than unity. It can be shown (see Lemma 3 in Section 6) that (9) is implied by the inequality $\|\hat{\mathbf{\Phi}}^2 P\| < 1$, which carries a useful interpretation. Because

$$\|\hat{\mathbf{\Phi}}^2 P\| = \max_{1 \le i \le K} \left\{ \sum_{k=1}^{K} \|\phi_k\|^2 p_{ik} \right\} = \max_{1 \le \mathbf{s}_{t-1} \le K} E(\|\phi_{\mathbf{s}_t}\|^2 | \mathbf{s}_{t-1}), \tag{10}$$

Assumption 2 is satisfied if the conditional expectation $E(\|\phi_{\mathbf{s}_t}\|^2 | \mathbf{s}_{t-1})$ is strictly less than unity for every state $\mathbf{s}_{t-1}$. Although stronger than (9), the condition $\|\hat{\mathbf{\Phi}}^2 P\| < 1$ is easier to verify and interpret.

Assumption 3. There are an integer $\nu \ge 1$ and a nonempty subset $\mathcal{J}_1 = \{j_1, \ldots, j_{K_1}\}$ of the state space $\mathcal{J} = \{1, \ldots, K\}$ such that

$$\min_{i \in \mathcal{J}, j \in \mathcal{J}_1} p_{ij}^{(\nu)} = \theta > 0, \tag{11}$$

where $p_{ij}^{(\nu)}$ is the (i,j)th element in $(P^\nu)'$.

In words, the preceding assumption requires that there is at least one state that can be reached (with a positive probability) in a finite number of steps from any state. Under this assumption, it follows from Doob (1953) that the stable unconditional probabilities

$$\pi_j = \mathcal{P}(\mathbf{s}_t = j), \qquad j = 1, \ldots, K \tag{12}$$

exist and can be expressed as

$$\pi_j = \lim_{\tau \to \infty} p_{ij}^{(\tau)}, \qquad j = 1, \ldots, K. \tag{13}$$

https://doi.org/10.1017/S026646660016102X Published online by Cambridge University Press

This in turn implies that the unconditional expectation of ξ_t is given by $E(\xi_t) = \pi$, where $\pi = [\pi_1, \ldots, \pi_K]'$.

We define a K-block diagonalization for any $n_1K \times n_2$ matrix M as

$$\operatorname{diag}_K(M) = \operatorname{diag}_K\left(\begin{bmatrix} M_1 \\ \vdots \\ M_K \end{bmatrix}\right) = \begin{bmatrix} M_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & M_K \end{bmatrix}_{n_1K \times n_2K},$$

where M_j's are $n_1 \times n_2$ matrices. For example, $\operatorname{diag}_K(J') = I_{nK \times nK}$. Further, for given matrices B_1 of size $(n_1K \times n_1K)$ and B_2 of size $(n_2K \times n_2K)$, we define an operator on any $n_1K \times n_2$ matrix M as

$$\mathcal{D}_{[B_1, B_2]}(M) = \operatorname{diag}_K(B_1M)B_2J'_{n_2}, \tag{14}$$

where $J_{n_2} = [I_{n_2}, \ldots, I_{n_2}]_{n_2 \times n_2K}$. Repeated operations are defined as

$$\mathcal{D}^i_{[B_1, B_2]}(M) = \mathcal{D}_{[B_1, B_2]}(\mathcal{D}^{i-1}_{[B_2, B_2]}(M))$$

for $i = 1, 2, \ldots$, where $\mathcal{D}^0_{[B_1, B_2]}(M) = M$. Note that both $\operatorname{diag}_K(\cdot)$ and $\mathcal{D}_{[B_1, B_2]}(\cdot)$ are linear in their argument, that is,

$$\operatorname{diag}_K(aM + bN) = a \operatorname{diag}_K(M) + b \operatorname{diag}_K(N)$$

and

$$\mathcal{D}_{[B_1, B_2]}(aM + bN) = a\mathcal{D}_{[B_1, B_2]}(M) + b\mathcal{D}_{[B_1, B_2]}(N)$$

hold for any scalars a, b and any matrices M, N of size $n_1K \times n_2$.

3. MAIN RESULTS

THEOREM 1. *Under Assumptions* 1 *and* 2, $\mathbf{x}_t$ *defined by* (3) *can be expressed as* (7) *and is uniformly bounded in* $\mathcal{L}_2$, *that is*, $\sup_t E(\mathbf{x}'_t\mathbf{x}_t) < \infty$.

The $\mathcal{L}_2$-boundedness of $\mathbf{x}_t$ does not directly rely on Assumption 3 or the existence of (12) and (13). Instead, it depends on the products of $\hat{\mathbf{\Phi}}$ and P. The condition for the $\mathcal{L}_2$-boundedness (and stationarity later on), given in terms of $\hat{\mathbf{\Phi}}^2P$ and $\hat{\mathbf{\Phi}}P$, generally differs from the stationarity condition for the conventional VAR model ($K = 1$) that requires $\mathbf{r}(\phi_1) < 1$. However, the interpretations for these conditions appear similar in the sense that both require a structure such that the impact of the past vanishes eventually. Given that Assumption 2 is implied by the condition $\|\hat{\mathbf{\Phi}}^2P\| < 1$ (see (10)), which requires the mean magnitude of the coefficient matrices to be less than unity, a trade-off between $\hat{\mathbf{\Phi}}$ and P is apparently possible. For instance, the $\mathcal{L}_2$-boundedness of $\mathbf{x}_t$ does not necessarily break down for the cases where $\|\phi_k\| \geq 1$ for some k (but not all k), provided that the corresponding p_{ik} is sufficiently small. Given that the radius of a matrix is no

https://doi.org/10.1017/S026646660016102X Published online by Cambridge University Press

greater than any of its norms, it is conceivable that the condition for the $\mathcal{L}_2$-boundedness could be weakened.

THEOREM 2. *Under Assumptions* 1 *and* 2 *and the existence of* (12), $\mathbf{x}_t$ *defined by* (3) *is covariance stationary and its first and second moments are given by*

$$\mu = E(\mathbf{x}_t) = J\bar{\mathbf{A}}, \tag{15}$$

$$\mathbf{\Gamma}_0 = \mathrm{Var}(\mathbf{x}_t) = J\bar{\mathbf{\Gamma}} - \mu\mu', \tag{16}$$

and

$$\mathbf{\Gamma}_\tau = \mathrm{Cov}(\mathbf{x}_{t+\tau}, \mathbf{x}_t) = J[\mathbf{\Phi}(P \otimes I_n)]^\tau \bar{\mathbf{\Gamma}} + \Delta_\tau - \mu\mu', \tag{17}$$

where $\tau > 0$,

$$\bar{\mathbf{A}} = E(\xi_t \otimes \mathbf{x}_t) = [I_{nK} - \mathbf{\Phi}(P \otimes I_n)]^{-1}\mathbf{A}\pi,$$

$$\bar{\mathbf{\Gamma}} = E[\xi_t \otimes (\mathbf{x}_t\mathbf{x}_t')] = \Sigma_0 + \Sigma_1 + \Sigma_2,$$

$$\Sigma_0 = \sum_{j=0}^{\infty} \mathcal{D}^j_{[\mathbf{\Phi}(P\otimes I_n),\mathbf{\Phi}']}(\bar{\mathbf{\Lambda}}(\pi \otimes I_n)),$$

$$\Sigma_1 = \sum_{j=0}^{\infty}\sum_{l=j+1}^{\infty} \mathcal{D}^j_{[\mathbf{\Phi}(P\otimes I_n),\mathbf{\Phi}']}(\mathrm{diag}_K\{[\mathbf{\Phi}(P \otimes I_n)]^{l-j}\mathbf{A}\pi\}\mathbf{A}'J'),$$

$$\Sigma_2 = \sum_{j=0}^{\infty}\sum_{l=j+1}^{\infty} \mathcal{D}^j_{[\mathbf{\Phi}(P\otimes I_n),\mathbf{\Phi}']}(\mathbf{A}\mathcal{D}^{l-j}_{[P,\mathbf{\Phi}']}(\mathrm{diag}_K(\pi)\mathbf{A}'J')),$$

$$\Delta_\tau = J\sum_{j=0}^{\tau-1} [\mathbf{\Phi}(P \otimes I_n)]^j \mathbf{A}P^{\tau-j}\,\mathrm{diag}_K\{\bar{\mathbf{A}}\}'J'.$$

We note that Assumption 3 implies the existence of (12). Unlike the conventional VAR process, the stationarity of $\{\mathbf{x}_t\}$ depends on the mixtures of $\hat{\mathbf{\Phi}}$ and P, which permits a certain trade-off between $\hat{\mathbf{\Phi}}$ and P. For instance, a large element in $\hat{\mathbf{\Phi}}$ will not destroy the stationarity of $\mathbf{x}_t$ provided that the corresponding elements in P are sufficiently small. This feature allows a stationary VAR-MS process to occasionally behave in a nonstationary (or explosive) fashion.

The mean of the VAR-MS process, $E(\mathbf{x}_t) = J[I_{nK} - \mathbf{\Phi}(P \otimes I_n)]^{-1}\mathbf{A}\pi$, is comparable with the mean for the conventional VAR process (VAR-MS with $K = 1$), $E(\mathbf{x}_t) = [I_n - \phi_1]^{-1}\alpha_1$. However, comparing the second moments reveals more dissimilarity than similarity. The main difference is that the conventional "constant term" in the VAR-MS model is no longer constant but serves as an additional innovation to the model. This results in the terms Δ_τ, Σ_1, and Σ_2, which would disappear if $\alpha_{\mathbf{s}_t} \equiv 0$.

THEOREM 3. *Under Assumptions* 1–3, *the autocovariance matrix* $\mathbf{\Gamma}_\tau$ *in* (17) *converges to zero at the rate* $O(\tau\delta^\tau)$ *as* $\tau \to \infty$, *where* $0 < \delta < 1$.

https://doi.org/10.1017/S026646660016102X Published online by Cambridge University Press

The decay rate of $\mathbf{\Gamma}_\tau$ slightly differs from the exponential rate of the autocovariance matrix for a stationary VAR process. The difference comes again from the fact that the term $\alpha_{\mathbf{s}_t}$ in the VAR-MS model functions as an innovation to the model. If $\alpha_{\mathbf{s}_t} \equiv 0$, the decay rate of $\mathbf{\Gamma}_\tau$ will be purely exponential (see the proof of Theorem 3). In view of the decay rate of $\mathbf{\Gamma}_\tau$, the stationary VAR-MS process has a "short memory" as opposed to the "long memory" of a stationary fractionally integrated ARMA process whose autocovariance decays at a hyperbolic rate (see Granger, 1980; and Hosking, 1981). Further, Theorem 3 immediately implies that $\bar{\mathbf{x}}_T \xrightarrow{\mathcal{L}_2} \mu$ and $\bar{\mathbf{x}}_T \xrightarrow{\text{a.s.}} \mu$ as $T \to \infty$, where $\bar{\mathbf{x}}_T = T^{-1} \sum_1^T \mathbf{x}_t$ (see, e.g., Davidson, 1994).

Because $\alpha_{\mathbf{s}_t}$ (or $\phi_{\mathbf{s}_t}$) acts as an innovation, it is appropriate to define a shock to $\mathbf{x}_t$ at t as the one-step optimal forecast error

$$\mathbf{w}_t = \mathbf{x}_t - E(\mathbf{x}_t | \mathcal{F}_{t-1}) = (\zeta_t' \otimes I_n)(\tilde{\alpha} + \tilde{\phi}\mathbf{x}_{t-1}) + \mathbf{u}_t, \tag{18}$$

where $\zeta_t = \xi_t - P\xi_{t-1}$. Whereas $\{\mathbf{w}_t\}$ is serially uncorrelated and stationary with zero mean and finite variance, its conditional variance (which is $\mathbf{x}_t$'s conditional variance also)

$$E(\mathbf{w}_t \mathbf{w}_t' | \mathcal{F}_{t-1}) = \sum_{j=1}^{K} p_{\mathbf{s}_{t-1}j}[\lambda_j + (\alpha_j + \phi_j \mathbf{x}_{t-1})(\alpha_j + \phi_j \mathbf{x}_{t-1})'] - \left[\sum_{j=1}^{K} p_{\mathbf{s}_{t-1}j}(\alpha_j + \phi_j \mathbf{x}_{t-1})\right]\left[\sum_{j=1}^{K} p_{\mathbf{s}_{t-1}j}(\alpha_j + \phi_j \mathbf{x}_{t-1})\right]' \tag{19}$$

is a function of $\mathbf{x}_{t-1}$ and ξ_{t-1}, which in turn depend on $\mathbf{w}_{t-1}$. In this sense, the VAR-MS model exhibits certain conditional heteroskedasticity similar to the ARCH effects introduced by Engle (1982).

THEOREM 4. *Under the conditions in Theorem* 3, $\mathbf{x}_t$ *can be decomposed as*

$$\mathbf{x}_t = \sum_{j=0}^{\infty} C_j \mathbf{w}_{t-j} + \mu, \tag{20}$$

where $\{\mathbf{w}_t\}$ *is defined in* (18), $C_0 = I_n$, *and* C_j *decays at the rate* $O(j\delta^j)$ *as* $j \to \infty$ *and* $0 < \delta < 1$.

It is noted that (20) is given in terms of the optimal prediction errors $\{\mathbf{w}_t\}$ and Wold's decomposition is usually given in terms of the errors of the linear projection of $\mathbf{x}_t$ on the space spanned by $\{\mathbf{x}_s : s \le t - 1\}$. Therefore, according to the uniqueness of Wold's decomposition, the optimal prediction and the linear projection of $\mathbf{x}_t$ agree. The impulse response of $\mathbf{x}_{t+\tau}$ to the shock $\mathbf{w}_t$ at t is characterized by the deterministic matrix C_τ, which converges to zero at the rate $O(\tau\delta^\tau)$. The process $\{\mathbf{x}_t\}$ is "mean reverting" in the sense that the impact of $\mathbf{w}_t$ on $\mathbf{x}_{t+\tau}$ dies out at the rate $O(\tau\delta^\tau)$.

https://doi.org/10.1017/S026646660016102X Published online by Cambridge University Press

4. AN EXAMPLE

We use a simple example to demonstrate the conditions that are required for stationarity. Consider the following data-generating process with $n = 1$ and $K = 2$:

$$x_t = \begin{cases} \phi_1 x_{t-1} + \varepsilon_t & \text{if } \mathbf{s}_t = 1 \\ \phi_2 x_{t-1} + \varepsilon_t & \text{if } \mathbf{s}_t = 2, \end{cases}$$

where $\{\varepsilon_t\}$ is i.i.d. $N(0, \sigma^2)$ and the transition probabilities of $\{\mathbf{s}_t\}$ are

$$p_{11} = \mathcal{P}(\mathbf{s}_t = 1 | \mathbf{s}_{t-1} = 1), \qquad p_{22} = \mathcal{P}(\mathbf{s}_t = 2 | \mathbf{s}_{t-1} = 2),$$

with $p_{12} = 1 - p_{11}$ and $p_{21} = 1 - p_{22}$. The simplicity of the data-generating process allows us to focus on the matrices

$$\mathbf{\Phi} = \begin{bmatrix} \phi_1 & 0 \\ 0 & \phi_2 \end{bmatrix}, \qquad \hat{\mathbf{\Phi}} = \begin{bmatrix} |\phi_1| & 0 \\ 0 & |\phi_2| \end{bmatrix}, \quad \text{and} \quad P = \begin{bmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{bmatrix},$$

which are the main concerns for stationarity.

Assumption 2, which ensures the boundedness of x_t, in this case is equivalent to

$$\max_{j=1,2} \Big\{ p_{11}|\phi_1|^j + p_{22}|\phi_2|^j + \sqrt{(p_{11}|\phi_1|^j + p_{22}|\phi_2|^j)^2 - 4|\phi_1\phi_2|^j(p_{11} + p_{22} - 1)} \Big\} < 2, \tag{21}$$

which defines a bounded region in the (ϕ_1, ϕ_2)-space for any $p_{11} \in (0,1)$ and $p_{22} \in (0,1)$. The trade-off between $\hat{\mathbf{\Phi}}$ and P can be seen from (21). For example, when $\phi_2 = 0$, the condition becomes $\phi_1^2 < 1/p_{11}$, allowing $|\phi_1| > 1$. As mentioned before, a stronger condition $\|\hat{\mathbf{\Phi}}^2 P\| < 1$, i.e.,

$$\max_{i=1,2} \{ p_{i1}|\phi_1|^2 + p_{i2}|\phi_2|^2 \} < 1 \tag{22}$$

may be used for boundedness and stationarity. The condition (22) defines the interior intersection of two ellipses in the (ϕ_1, ϕ_2)-space. The regions defined by (21) and (22) are illustrated in Figure 1 with $p_{11} = 0.8$ and $p_{22} = 0.7$. Because of symmetry, only the first quadrant of the (ϕ_1, ϕ_2)-space is portrayed. It can be noted from Figure 1 that the stationary region (21) is given by the interior of the solid curve ($j = 2$) and two axes and is, as expected, larger than the region defined by (22). Clearly, the stationary region includes the areas where one of $|\phi_1|$ and $|\phi_2|$ is greater or equal to unity.

In addition to Assumption 2, the existence of (12) is required for x_t to be covariance stationary. This and the requirements in Theorem 3 are fulfilled if either $p_{11} \in (0,1)$ δ $p_{22} \in (0,1)$ (see the remark following Assumption 3). There are four extreme cases where p_{11} and p_{22} are either 0 or 1. When $(p_{11}, p_{22}) = (0,1)$ or $(1,0)$, x_t collapses to a single AR(1) process. The case with $(p_{11}, p_{22}) = (1,1)$ gives rise to two AR(1) regimes that never interact (with probability 1), and the behavior of

https://doi.org/10.1017/S026646660016102X Published online by Cambridge University Press

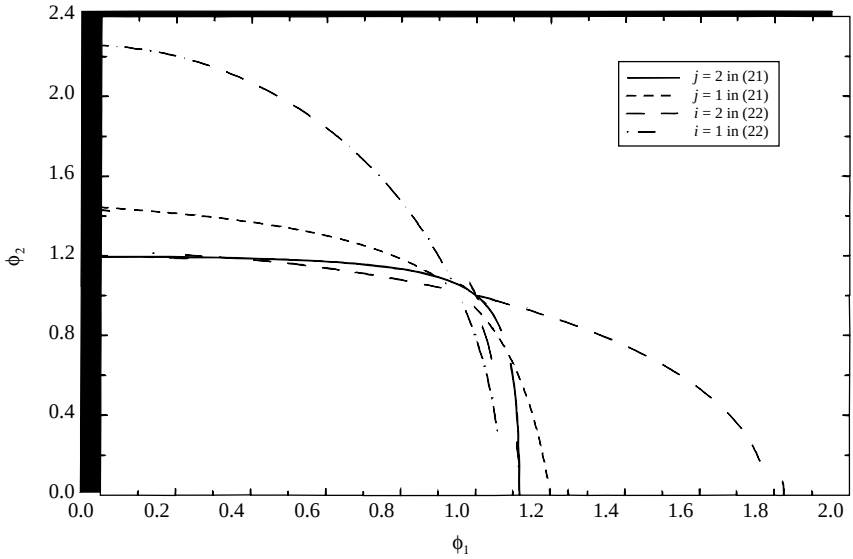


FIGURE 1. Stationary region ($p_{11} = 0.8, p_{22} = 0.7$).

x_t depends on the starting state. In the case of $(p_{11}, p_{22}) = (0,0)$, the Markov chain is periodic, and x_t switches regimes every period (with probability 1). For the latter two cases, as the stable unconditional distribution (12) of the Markov chain does not exist, the results in Theorems 2–4 are no longer guaranteed even when Assumption 2 holds.

Under the stationarity conditions, Theorem 2 indicates that the mean is $E(x_t) = 0$, and the autocovariance becomes

$$\mathrm{Cov}(x_{t+\tau}, x_t) = \mathbf{1}(\mathbf{\Phi}P)^\tau \sum_{j=0}^{\infty} \mathcal{D}^j_{[\mathbf{\Phi}P, \mathbf{\Phi}']}(\pi)\sigma^2 = \mathbf{1}(\mathbf{\Phi}P)^\tau (I_2 - \mathbf{\Phi}^2 P)^{-1}\pi\sigma^2,$$

where $\pi = [\pi_1, \pi_2]'$ contains the stable unconditional probabilities of the Markov chain $\pi_1 = \mathcal{P}(\mathbf{s}_t = 1)$ and $\pi_2 = \mathcal{P}(\mathbf{s}_t = 2)$. Note that the operator $\mathcal{D}_{[\mathbf{\Phi}P, \mathbf{\Phi}']}$ is reduced to the matrix multiplication $\mathcal{D}_{[\mathbf{\Phi}P, \mathbf{\Phi}']}(\cdot) = \mathbf{\Phi}^2 P(\cdot)$ when $n = 1$. Further, according to (19), the conditional variance of x_t is given by

$$\begin{aligned}\mathrm{Var}(x_t|\mathcal{F}_{t-1}) &= \sigma^2 + \sum_{j=1}^{2} p_{\mathbf{s}_{t-1}j}(\phi_j x_{t-1})^2 - \left[\sum_{j=1}^{2} p_{\mathbf{s}_{t-1}j}(\phi_j x_{t-1})\right]^2 \\ &= \sigma^2 + \mathrm{Var}(\phi_{\mathbf{s}_t}|\mathbf{s}_{t-1})x_{t-1}^2,\end{aligned}$$

which is clearly a function of x_{t-1} and $\mathbf{s}_{t-1}$ and therefore time varying. It is similar to the ARCH(1) specification of Engle (1982). In fact, for the simple (Bernoulli)

https://doi.org/10.1017/S026646660016102X Published online by Cambridge University Press

switching model where $p_{22} = 1 - p_{11}$, $\mathrm{Var}(\phi_{\mathbf{s}_t}|\mathbf{s}_{t-1}) = \mathrm{Var}(\phi_{\mathbf{s}_t})$ is a constant, and the conditional variance of x_t is exactly an ARCH(1) process.

5. CONCLUSION

Under the stationarity conditions given in the paper, the long-run behavior of the VAR-MS process turns out to be similar to that of the stationary VAR process. The strength of the VAR-MS model lies in its short-run dynamics that conventional linear processes are unable to offer. A particularly interesting feature is that the VAR-MS process can occasionally behave in a nonstationary manner although being stationary and mean reverting in the long run.

6. PROOFS

LEMMA 1. *For any t, denote* $\psi_t(l,j) = \phi(t,l)\mathbf{v}_{t-l}\mathbf{v}'_{t-j}\phi(t,j)'$, *where* $j,l = 1,2,\ldots$. *Then*

$$E(\xi_t \otimes \mathbf{v}_t|\mathcal{F}_{t-i}) = \mathbf{A}P^i\xi_{t-i} \tag{23}$$

holds for $i = 1,2,\ldots$;

$$E(\xi_t \otimes \phi(t,j)|\mathcal{F}_{t-i}) = [\mathbf{\Phi}(P \otimes I_n)]^i[\xi_{t-i} \otimes \phi(t-i,j-i)] \tag{24}$$

and

$$\begin{aligned} &E[\xi_t\mathbf{v}'_{t-j}\phi(t,j)'|\mathcal{F}_{t-i}] \\ &\quad = \underbrace{\mathrm{diag}_K[P\,\mathrm{diag}_K[P\ldots\,\mathrm{diag}_K[P}_{i\text{ folds}}(\xi_{t-i}\mathbf{v}'_{t-j}\phi(t-i,j-i)')]\underbrace{\mathbf{\Phi}'J'\ldots]\mathbf{\Phi}'J']\mathbf{\Phi}'J'}_{i\text{ folds}} \\ &\quad = \mathcal{D}^i_{[P,\mathbf{\Phi}']}(\xi_{t-i}\mathbf{v}'_{t-j}\phi(t-i,j-i)') \end{aligned} \tag{25}$$

hold for $i = 1,\ldots,j$;

$$\begin{aligned} &E[\xi_t \otimes \psi_t(l,j)|\mathcal{F}_{t-i}] \\ &\quad = \underbrace{\mathrm{diag}_K[\mathbf{\Phi}(P \otimes I_n)\mathrm{diag}_K[\mathbf{\Phi}(P \otimes I_n)\ldots\mathrm{diag}_K[\mathbf{\Phi}(P \otimes i_n)}_{i\text{ folds}} \\ &\qquad (\xi_{t-i} \otimes \psi_{t-i}(l-i,j-i))]\underbrace{\mathbf{\Phi}'J'\ldots]\mathbf{\Phi}'J']\mathbf{\Phi}'J'}_{i\text{ folds}} \\ &\quad = \mathcal{D}^i_{[\mathbf{\Phi}(P\otimes I_n),\mathbf{\Phi}']}(\xi_{t-i} \otimes \psi_{t-i}(l-i,j-i)) \end{aligned} \tag{26}$$

https://doi.org/10.1017/S026646660016102X Published online by Cambridge University Press

and

$$E[\xi_t\|\phi(t,l)\mathbf{v}_{t-l}\|\|\phi(t,j)\mathbf{v}_{t-j}\||\mathcal{F}_{t-i}]$$
$$\le(\hat{\mathbf{\Phi}}^2P)^i\xi_{t-i}\|\phi(t-i,l-i)\mathbf{v}_{t-l}\|\|\phi(t-i,j-i)\mathbf{v}_{t-j}\| \quad (27)$$

hold for $i=1,\dots,\min(j,l)$.

Proof. This lemma can be inductively proved. Because of similarity, we only prove the inequality (27). First, it is true for $i=1$ as

$$E(\xi_t\|\phi(t,l)\mathbf{v}_{t-l}\|\|\phi(t,j)\mathbf{v}_{t-j}\||\mathcal{F}_{t-1})$$
$$\le\sum_{k=1}^{K}\mathbf{e}_k\|\phi_k\|^2p_{\mathbf{s}_{t-1}k}\|\phi(t-1,l-1)\mathbf{v}_{t-l}\|\|\phi(t-1,j-1)\mathbf{v}_{t-j}\|$$
$$=\begin{bmatrix}\|\phi_1\|^2p_{\mathbf{s}_{t-1}1}\\ \vdots\\ \|\phi_K\|^2p_{\mathbf{s}_{t-1}K}\end{bmatrix}\|\phi(t-1,l-1)\mathbf{v}_{t-l}\|\|\phi(t-1,j-1)\mathbf{v}_{t-j}\|$$
$$=\hat{\mathbf{\Phi}}^2P\xi_{t-1}\|\phi(t-1,l-1)\mathbf{v}_{t-l}\|\|\phi(t-1,j-1)\mathbf{v}_{t-j}\|.$$

On assuming that it holds for $i-1$,

$$E(\xi_t\|\phi(t,l)\mathbf{v}_{t-l}\|\|\phi(t,j)\mathbf{v}_{t-j}\||\mathcal{F}_{t-i})$$
$$\le(\hat{\mathbf{\Phi}}^2P)^{i-1}E(\xi_{t-i+1}\|\phi(t-i+1,l-i+1)\mathbf{v}_{t-l}\|$$
$$\times\|\phi(t-i+1,j-i+1)\mathbf{v}_{t-j}\||\mathcal{F}_{t-i})$$
$$\le(\hat{\mathbf{\Phi}}^2P)^{i-1}\sum_{k=1}^{K}\mathbf{e}_k\|\phi_k\|^2p_{\mathbf{s}_{t-i}k}\|\phi(t-i,l-i)\mathbf{v}_{t-l}\|\|\phi(t-i,j-i)\mathbf{v}_{t-j}\|$$
$$=(\hat{\mathbf{\Phi}}^2P)^{i-1}\begin{bmatrix}\|\phi_1\|^2p_{\mathbf{s}_{t-i}1}\\ \vdots\\ \|\phi_K\|^2p_{\mathbf{s}_{t-i}K}\end{bmatrix}\|\phi(t-i,l-i)\mathbf{v}_{t-l}\|\|\phi(t-i,j-i)\mathbf{v}_{t-j}\|$$
$$=(\hat{\mathbf{\Phi}}^2P)^{i-1}\hat{\mathbf{\Phi}}^2P\xi_{t-i}\|\phi(t-i,l-i)\mathbf{v}_{t-l}\|\|\phi(t-i,j-i)\mathbf{v}_{t-j}\|$$

shows (27) is also true for the case of i. Therefore it holds for all $i=1,\dots,\min\{j,l\}$ according to the induction principle. ■

LEMMA 2. *For any* t *and* $j=1,2,\dots,$

$$E[\phi(t,j)\mathbf{u}_{t-j}]=0. \quad (28)$$

Further, for $l\ge j$

$$E[\|\phi(t,l)\mathbf{v}_{t-l}\|\|\phi(t,j)\mathbf{v}_{t-j}\|]\le n\mathbf{1}(\hat{\mathbf{\Phi}}^2P)^j(\hat{\mathbf{\Phi}}P)^{l-j}\mathbf{1}'\max_{1\le k\le K}\{\mathrm{tr}(\bar{\lambda}_k)\}, \quad (29)$$

where $\bar{\lambda}_k=\lambda_k+\alpha_k\alpha_k'$.

https://doi.org/10.1017/S026646660016102X Published online by Cambridge University Press

Proof. Noting the fact that $\phi(t,j) = J(\xi_t \otimes \phi(t,j))$ and $\phi(t,0) = I_n$ and setting $i = j$ in (24), we find

$$E[\phi(t,j)\mathbf{u}_{t-j}|\mathcal{F}_{t-j}] = J[\mathbf{\Phi}(P \otimes I_n)]^j[\xi_{t-j} \otimes \mathbf{u}_{t-j}].$$

According to (4),

$$E[\phi(t,j)\mathbf{u}_{t-j}|\mathbf{s}_{t-j},\mathcal{F}_{t-j-1}] = J[\mathbf{\Phi}(P \otimes I_n)]^j[\xi_{t-j} \otimes E(\mathbf{u}_{t-j}|\mathbf{s}_{t-j},\mathcal{F}_{t-j-})] = 0,$$

implying (28). To show (29), we note the fact that

$$\|\phi(t,l)\mathbf{v}_{t-l}\|\|\phi(t,j)\mathbf{v}_{t-j}\| = \mathbf{1}\xi_t\|\phi(t,l)\mathbf{v}_{t-l}\|\|\phi(t,j)\mathbf{v}_{t-j}\|.$$

Then, setting $i = j$, (27) implies

$$E(\|\phi(t,l)\mathbf{v}_{t-l}\|\|\phi(t,j)\mathbf{v}_{t-j}\||\mathcal{F}_{t-j}) \le \mathbf{1}(\hat{\mathbf{\Phi}}^2P)^j\|\mathbf{v}_{t-j}\|\xi_{t-j}\|\phi(t-j,l-j)\mathbf{v}_{t-l}\|.$$

It follows that

$$\begin{aligned}
&E(\|\phi(t,l)\mathbf{v}_{t-l}\|\|\phi(t,j)\mathbf{v}_{t-j}\||\mathbf{s}_{t-j},\mathcal{F}_{t-j-1}) \\
&\quad\le \mathbf{1}(\hat{\mathbf{\Phi}}^2P)^jE[\|\mathbf{v}_{t-j}\||\mathbf{s}_{t-j},\mathcal{F}_{t-j-1}]\xi_{t-j}\|\phi(t-j,l-j)\mathbf{v}_{t-l}\| \\
&\quad\le \mathbf{1}(\hat{\mathbf{\Phi}}^2P)^jn^{1/2}E[(\mathbf{v}'_{t-j}\mathbf{v}_{t-j})^{1/2}|\mathbf{s}_{t-j},\mathcal{F}_{t-j-1}]\xi_{t-j}\|\phi(t-j,l-j)\mathbf{v}_{t-l}\| \\
&\quad\le n^{1/2}\mathbf{1}(\hat{\mathbf{\Phi}}^2P)^jE[\mathbf{v}'_{t-j}\mathbf{v}_{t-j}|\mathbf{s}_{t-j},\mathcal{F}_{t-j-1}]^{1/2}\xi_{t-j}\|\phi(t-j,l-j)\mathbf{v}_{t-l}\| \\
&\quad\le n^{1/2}\mathbf{1}(\hat{\mathbf{\Phi}}^2P)^j\mathrm{tr}(\bar{\lambda}_{\mathbf{s}_{t-j}})^{1/2}\xi_{t-j}\|\phi(t-j,l-j)\mathbf{v}_{t-l}\| \\
&\quad\le n^{1/2}\mathbf{1}(\hat{\mathbf{\Phi}}^2P)^j\max_{1\le k\le K}\{\mathrm{tr}(\bar{\lambda}_k)\}^{1/2}\xi_{t-j}\|\phi(t-j,l-j)\mathbf{v}_{t-l}\|, \qquad \textbf{(30)}
\end{aligned}$$

where the second inequality follows the fact that $\|\mathbf{v}_t\| \le n^{1/2}(\mathbf{v}'_t\mathbf{v}_t)^{1/2}$ and the third is implied by Jesen's inequality. Similar to (27), it can be shown that

$$E[\xi_{t-j}\|\phi(t,l)\mathbf{v}_{t-l}\||\mathcal{F}_{t-i}] \le (\hat{\mathbf{\Phi}}P)^i\xi_{t-i}\|\phi(t-i,l-i)\mathbf{v}_{t-l}\|$$

for $i = 1,\ldots,l$. This, together with (30), gives

$$\begin{aligned}
&E(\|\phi(t,l)\mathbf{v}_{t-l}\|\|\phi(t,j)\mathbf{v}_{t-j}\||\mathcal{F}_{t-l}) \\
&\quad\le n^{1/2}\mathbf{1}(\hat{\mathbf{\Phi}}^2P)^j\max_{1\le k\le K}\{\mathrm{tr}(\bar{\lambda}_k)\}^{1/2}(\hat{\mathbf{\Phi}}P)^{l-j}\xi_{t-l}\|\mathbf{v}_{t-l}\| \\
&\quad\le n^{1/2}\mathbf{1}(\hat{\mathbf{\Phi}}^2P)^j\max_{1\le k\le K}\{\mathrm{tr}(\bar{\lambda}_k)\}^{1/2}(\hat{\mathbf{\Phi}}P)^{l-j}\mathbf{1}'\|\mathbf{v}_{t-l}\| \\
&\quad\le n\mathbf{1}(\hat{\mathbf{\Phi}}^2P)^j\max_{1\le k\le K}\{\mathrm{tr}(\bar{\lambda}_k)\}^{1/2}(\hat{\mathbf{\Phi}}P)^{l-j}\mathbf{1}'(\mathbf{v}'_{t-l}\mathbf{v}_{t-l})^{1/2}
\end{aligned}$$

https://doi.org/10.1017/S026646660016102X Published online by Cambridge University Press

and

$$
\begin{aligned}
&E(\|\phi(t,l)\mathbf{v}_{t-l}\|\|\phi(t,j)\mathbf{v}_{t-j}\|\,|\,\mathbf{s}_{t-l},\mathcal{F}_{t-l-1})\\
&\quad\le n\mathbf{1}(\hat{\mathbf{\Phi}}^2P)^j \max_{1\le k\le K}\{\mathrm{tr}(\bar{\lambda}_k)\}^{1/2}(\hat{\mathbf{\Phi}}P)^{l-j}\mathbf{1}' \max_{1\le k\le K}\{\mathrm{tr}(\bar{\lambda}_k)\}^{1/2}\\
&\quad= n\mathbf{1}(\hat{\mathbf{\Phi}}^2P)^j(\hat{\mathbf{\Phi}}P)^{l-j}\mathbf{1}' \max_{1\le k\le K}\{\mathrm{tr}(\bar{\lambda}_k)\}.
\end{aligned}
$$

The proof is completed by taking expectations on both sides of the preceding inequality. Strictly speaking, we have only proved (29) for the cases of $l > j$. However, when $l = j$, (30) reduces to

$$
\begin{aligned}
&E(\|\phi(t,l)\mathbf{v}_{t-l}\|\|\phi(t,j)\mathbf{v}_{t-j}\|\,|\,\mathbf{s}_{t-j},\mathcal{F}_{t-j-1})\\
&\quad\le \mathbf{1}(\hat{\mathbf{\Phi}}^2P)^jE[\|\mathbf{v}_{t-j}\|^2\,|\,\mathbf{s}_{t-j},\mathcal{F}_{t-j-1}]\xi_{t-j}\\
&\quad\le n\mathbf{1}(\hat{\mathbf{\Phi}}^2P)^j\mathbf{1}'E[\mathbf{v}_{t-j}'\mathbf{v}_{t-j}\,|\,\mathbf{s}_{t-j},\mathcal{F}_{t-j-1}]\\
&\quad\le n\mathbf{1}(\hat{\mathbf{\Phi}}^2P)^j\mathbf{1}'\,\mathrm{tr}(\bar{\lambda}_{\mathbf{s}_{t-j}})\\
&\quad\le n\mathbf{1}(\hat{\mathbf{\Phi}}^2P)^j\mathbf{1}' \max_{1\le k\le K}\{\mathrm{tr}(\bar{\lambda}_k)\},
\end{aligned}
$$

which implies (29) in the case of $l = j$. ■

LEMMA 3.

(i) $\|\hat{\mathbf{\Phi}}P\|^2 \le \|\hat{\mathbf{\Phi}}^2P\|$;
(ii) $\|\mathbf{\Phi}(P\otimes I_n)\| = \|\hat{\mathbf{\Phi}}P\|$.

Proof. Denote $\varphi(i) = \sum_{k=1}^K \|\phi_k\|p_{ik}$. Then

$$
\sum_{k=1}^K(\|\phi_k\| - \varphi(i))^2p_{ik} = \sum_{k=1}^K\|\phi_k\|^2p_{ik} - \varphi(i)^2 \ge 0
$$

implies $\varphi(i)^2 \le \sum_{k=1}^K\|\phi_k\|^2p_{ik}$ for every i. Therefore

$$
\|\hat{\mathbf{\Phi}}P\|^2 = \max_{1\le i\le K}\varphi(i)^2 \le \max_{1\le i\le K}\sum_{k=1}^K\|\phi_k\|^2p_{ik} = \|\hat{\mathbf{\Phi}}^2P\|,
$$

as required in (i). Further, let $\phi_k^{(lj)}$ be the (l,j)th element in ϕ_k. We find

$$
\begin{aligned}
\|\mathbf{\Phi}(P\otimes I_n)\| &= \max_{1\le i\le K;1\le j\le n}\sum_{k=1}^K\sum_{l=1}^n|\phi_k^{(lj)}|p_{ik}\\
&= \max_{1\le i\le K}\sum_{k=1}^K\left(\max_{1\le j\le n}\sum_{l=1}^n|\phi_k^{(lj)}|\right)p_{ik}\\
&= \max_{1\le i\le K}\sum_{k=1}^K(\|\phi_k\|)p_{ik} = \|\hat{\mathbf{\Phi}}P\|,
\end{aligned}
$$

https://doi.org/10.1017/S026646660016102X Published online by Cambridge University Press

as required in (ii). Note that $\|\hat{\mathbf{\Phi}}^2 P\| < 1$ implies $\|\hat{\mathbf{\Phi}} P\| < 1$ and therefore Assumption 2. ■

Proof of Theorem 1. To show $\sum_{j=0}^{\infty} \phi(t,j)\mathbf{v}_{t-j} \in \mathcal{L}_2$, it suffices to show

$$E\left\| \sum_{j=m_1}^{m_2} \phi(t,j)\mathbf{v}_{t-j} \right\|^2 \to 0$$

as $m_1 \to \infty$ for any $m_2 > m_1$. First, we note that

$$\left\| \sum_{j=m_1}^{m_2} \phi(t,j)\mathbf{v}_{t-j} \right\|^2$$

$$\leq \sum_{j=m_1}^{m_2} \|\phi(t,j)\mathbf{v}_{t-j}\|^2 + 2 \sum_{j=m_1}^{m_2-1} \sum_{l=j+1}^{m_2} \|\phi(t,l)\mathbf{v}_{t-l}\| \|\phi(t,j)\mathbf{v}_{t-j}\|.$$

Applying (29) to the preceding inequality, we find

$$\frac{1}{n} E\left\| \sum_{j=m_1}^{m_2} \phi(t,j)\mathbf{v}_{t-j} \right\|^2$$

$$\leq \sum_{j=m_1}^{m_2} \mathbf{1}(\hat{\mathbf{\Phi}}^2 P)^j \mathbf{1}' \max_{1\leq k\leq K} \{\mathrm{tr}(\bar{\lambda}_k)\}$$

$$+ 2 \sum_{j=m_1}^{m_2-1} \sum_{l=j+1}^{m_2} \mathbf{1}(\hat{\mathbf{\Phi}}^2 P)^j (\hat{\mathbf{\Phi}} P)^{l-j} \mathbf{1}' \max_{1\leq k\leq K} \{\mathrm{tr}(\bar{\lambda}_k)\}$$

$$\leq \sum_{j=m_1}^{\infty} \mathbf{1}(\hat{\mathbf{\Phi}}^2 P)^j \mathbf{1}' \max_{1\leq k\leq K} \{\mathrm{tr}(\bar{\lambda}_k)\}$$

$$+ 2 \sum_{j=m_1}^{\infty} \sum_{l=j+1}^{\infty} \mathbf{1}(\hat{\mathbf{\Phi}}^2 P)^j (\hat{\mathbf{\Phi}} P)^{l-j} \mathbf{1}' \max_{1\leq k\leq K} \{\mathrm{tr}(\bar{\lambda}_k)\}$$

$$= \mathbf{1}(\hat{\mathbf{\Phi}}^2 P)^{m_1} \left(\sum_{j=m_1}^{\infty} (\hat{\mathbf{\Phi}}^2 P)^{j-m_1} + 2 \sum_{j=m_1}^{\infty} (\hat{\mathbf{\Phi}}^2 P)^{j-m_1} \sum_{l=j+1}^{\infty} (\hat{\mathbf{\Phi}} P)^{l-j} \right) \mathbf{1}'$$

$$\times \max_{1\leq k\leq K} \{\mathrm{tr}(\bar{\lambda}_k)\}.$$

Under the stated condition,

$$\sum_{j=0}^{\infty} (\hat{\mathbf{\Phi}}^2 P)^j = (I_K - \hat{\mathbf{\Phi}}^2 P)^{-1} \quad \text{and} \quad \sum_{j=0}^{\infty} (\hat{\mathbf{\Phi}} P)^j = (I_K - \hat{\mathbf{\Phi}} P)^{-1}.$$

https://doi.org/10.1017/S026646660016102X Published online by Cambridge University Press

Therefore the preceding inequality becomes

$$\begin{aligned}\frac{1}{n}E\left\|\sum_{j=m_1}^{m_2}\phi(t,j)\mathbf{v}_{t-j}\right\|^2 &\le \mathbf{1}(\hat{\mathbf{\Phi}}^2P)^{m_1}\left(\sum_{j=0}^{\infty}(\hat{\mathbf{\Phi}}^2P)^j+2\sum_{j=0}^{\infty}(\hat{\mathbf{\Phi}}^2P)^j(\hat{\mathbf{\Phi}}P)(I_K-\hat{\mathbf{\Phi}}P)^{-1}\right)\mathbf{1}'\\ &\quad\times\max_{1\le k\le K}\{\mathrm{tr}(\bar{\lambda}_k)\}\\ &=\mathbf{1}(\hat{\mathbf{\Phi}}^2P)^{m_1}((I_K-\hat{\mathbf{\Phi}}^2P)^{-1}[I_K+2(\hat{\mathbf{\Phi}}P)(I_K-\hat{\mathbf{\Phi}}P)^{-1}])\mathbf{1}'\max_{1\le k\le K}\{\mathrm{tr}(\bar{\lambda}_k)\},\end{aligned} \tag{31}$$

concluding that $E\|\sum_{j=m_1}^{m_2}\phi(t,j)\mathbf{v}_{t-j}\|^2\to 0$ as $m_1\to\infty$. ■

LEMMA 4. *For any t, denote $\psi_t(l,j)=\phi(t,l)\mathbf{v}_{t-l}\mathbf{v}_{t-j}'\phi(t,j)'$, where $j,l=1,2,\ldots$. If the unconditional probabilities $\pi_j=\mathcal{P}(\mathbf{s}_t=j)(1\le j\le K)$ exist for the Markov chain $\{\mathbf{s}_t\}$, then*

$$E[\xi_t\otimes\phi(t,j)\mathbf{v}_{t-j}]=[\mathbf{\Phi}(P\otimes I_n)]^j\mathbf{A}\pi, \tag{32}$$

$$E[\xi_t\otimes(\phi(t,j)\mathbf{v}_{t-j}\xi_t')]=\mathrm{diag}_K\{[\mathbf{\Phi}(P\otimes I_n)]^j\mathbf{A}\pi\}, \tag{33}$$

and

$$E[\xi_t\otimes\psi_t(j,j)]=\mathcal{D}^j_{[\mathbf{\Phi}(P\otimes I_n),\mathbf{\Phi}']}(\bar{\mathbf{\Lambda}}(\pi\otimes I_n)) \tag{34}$$

hold for $j=1,2,\ldots$;

$$E[\xi_t\otimes\psi_t(l,j)]=\mathcal{D}^j_{[\mathbf{\Phi}(P\otimes I_n),\mathbf{\Phi}']}(\mathrm{diag}_K\{[\mathbf{\Phi}(P\otimes I_n)]^{l-j}\mathbf{A}\pi\}\mathbf{A}'J') \tag{35}$$

and

$$E[\xi_t\otimes\psi_t(l,j)']=\mathcal{D}^j_{[\mathbf{\Phi}(P\otimes I_n),\mathbf{\Phi}']}(\mathbf{A}\mathcal{D}^{l-j}_{[P,\mathbf{\Phi}]}(\mathrm{diag}_K(\pi)\mathbf{A}'J')) \tag{36}$$

hold for $j<l$.

Proof. Using (24), we find

$$E[\xi_t\otimes\phi(t,j)\mathbf{v}_{t-j}|\mathcal{F}_{t-j}]=[\mathbf{\Phi}(P\otimes I_n)]^j[\xi_{t-j}\otimes\mathbf{v}_{t-j}].$$

Then, (28) and Assumption 3 imply

$$\begin{aligned}E[\xi_t\otimes\phi(t,j)\mathbf{v}_{t-j}]&=[\mathbf{\Phi}(P\otimes I_n)]^j\sum_{k=1}^{K}[\mathbf{e}_k\otimes\alpha_k]\pi_k\\ &=[\mathbf{\Phi}(P\otimes I_n)]^j\begin{bmatrix}\alpha_1\pi_1\\ \vdots\\ \alpha_K\pi_K\end{bmatrix}\\ &=[\mathbf{\Phi}(P\otimes I_n)]^j\mathbf{A}\pi,\end{aligned}$$

https://doi.org/10.1017/S026646660016102X Published online by Cambridge University Press

as required in (32). To show (33), we notice

$$
\begin{aligned}
&E[\xi_t \otimes (\phi(t,j)\mathbf{v}_{t-j}\xi_t')|\mathcal{F}_{t-1}] \\
&\quad = \sum_{k=1}^{K} [\mathbf{e}_k \otimes (\phi_k \phi(t-1,j-1)\mathbf{v}_{t-j}\mathbf{e}_k' p_{\mathbf{s}_{t-1}k})] \\
&\quad = \begin{bmatrix} \phi_1 p_{\mathbf{s}_{t-1}k}\phi(t-1,j-1)\mathbf{v}_{t-j}\mathbf{e}_1' \\ \vdots \\ \phi_K p_{\mathbf{s}_{t-1}K}\phi(t-1,j-1)\mathbf{v}_{t-j}\mathbf{e}_K' \end{bmatrix} \\
&\quad = \begin{bmatrix} \phi_1 p_{\mathbf{s}_{t-1}k}\phi(t-1,j-1)\mathbf{v}_{t-j} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \phi_K p_{\mathbf{s}_{t-1}K}\phi(t-1,j-1)\mathbf{v}_{t-j} \end{bmatrix} \\
&\quad = \mathrm{diag}_K\{[\mathbf{\Phi}(P \otimes I_n)](\xi_{t-1} \otimes \phi(t-1,j-1)\mathbf{v}_{t-j})\},
\end{aligned}
$$

which, together with (32), implies (33). To prove (34) and (35), we set $i = j$ in (26) and find

$$
E[\xi_t \otimes \psi_t(l,j)|\mathcal{F}_{t-j}] = \mathcal{D}^j_{[\mathbf{\Phi}(P\otimes I_n),\mathbf{\Phi}']}(\xi_{t-j} \otimes (\phi(t-j,l-j)\mathbf{v}_{t-l}\mathbf{v}_{t-j}')) \qquad \textbf{(37)}
$$

for $j \le l$. When $j = l$, it follows that

$$
\begin{aligned}
E[\xi_t \otimes \psi_t(j,j)|\mathbf{s}_{t-j},\mathcal{F}_{t-j-1}] &= \mathcal{D}^j_{[\mathbf{\Phi}(P\otimes I_n),\mathbf{\Phi}']}(\xi_{t-j} \otimes E[\mathbf{v}_{t-j}\mathbf{v}_{t-j}'|\mathbf{s}_{t-j},\mathcal{F}_{t-j-1}]) \\
&= \mathcal{D}^j_{[\mathbf{\Phi}(P\otimes I_n),\mathbf{\Phi}']}(\xi_{t-j} \otimes \bar{\lambda}_{\mathbf{s}_{t-j}}),
\end{aligned}
$$

and, from the existence $\pi_j = \mathcal{P}(\mathbf{s}_t = j)$ for j,

$$
\begin{aligned}
E[\xi_t \otimes \psi_t(j,j)] &= \mathcal{D}^j_{[\mathbf{\Phi}(P\otimes I_n),\mathbf{\Phi}']}\left(\sum_{k=1}^{K}(\mathbf{e}_k \otimes \bar{\lambda}_k)\pi_k\right) \\
&= \mathcal{D}^j_{[\mathbf{\Phi}(P\otimes I_n),\mathbf{\Phi}']}\left(\begin{bmatrix} \bar{\lambda}_1\pi_1 \\ \vdots \\ \bar{\lambda}_K\pi_K \end{bmatrix}\right),
\end{aligned}
$$

as required in (34). When $j < l$, (37) implies

$$
\begin{aligned}
&E[\xi_t \otimes \psi_t(l,j)|\mathcal{F}_{t-j-1}] \\
&\quad = \mathcal{D}^j_{[\mathbf{\Phi}(P\otimes I_n),\mathbf{\Phi}']}\left(\sum_{k=1}^{K}\mathbf{e}_k \otimes (\phi_k\phi(t-j-1,l-j-1)\mathbf{v}_{t-l}\alpha_k')p_{\mathbf{s}_{t-j-1}k}\right) \\
&\quad = \mathcal{D}^j_{[\mathbf{\Phi}(P\otimes I_n),\mathbf{\Phi}']}\left(\begin{bmatrix} \phi_1 p_{\mathbf{s}_{t-j-1},1}\phi(t-j-1,l-j-1)\mathbf{v}_{t-l}\alpha_1' \\ \vdots \\ \phi_K p_{\mathbf{s}_{t-j-1},K}\phi(t-j-1,l-j-1)\mathbf{v}_{t-l}\alpha_K' \end{bmatrix}\right) \\
&\quad = \mathcal{D}^j_{[\mathbf{\Phi}(P\otimes I_n),\mathbf{\Phi}']} \\
&\qquad \times (\mathrm{diag}_K\{\mathbf{\Phi}(P \otimes I_n)[\xi_{t-j-1} \otimes \phi(t-j-1,l-j-1)\mathbf{v}_{t-l}]\}\mathbf{A}'J').
\end{aligned}
$$

https://doi.org/10.1017/S026646660016102X Published online by Cambridge University Press

This, together with (24) and (32), delivers (35). For the case of (36), setting $i = j$ in (26) gives

$$E[\xi_t \otimes \psi_t(l,j)' \,|\, \mathcal{F}_{t-j}] = \mathcal{D}^j_{[\mathbf{\Phi}(P\otimes I_n),\mathbf{\Phi}']}(\xi_{t-j} \otimes (\mathbf{v}_{t-j}\mathbf{v}'_{t-l}\phi(t-j,l-j)')),$$

which implies

$$\begin{aligned}
&E[\xi_t \otimes \psi_t(l,j)' \,|\, \mathcal{F}_{t-j-1}] \\
&\quad = \mathcal{D}^j_{[\mathbf{\Phi}(P\otimes I_n),\mathbf{\Phi}']}\left(\sum_{k=1}^{K} \mathbf{e}_k \otimes (\alpha_k \mathbf{v}'_{t-l}\phi(t-j-1,l-j-1)'\phi'_k)p_{\mathbf{s}_{t-j-1}k}\right) \\
&\quad = \mathcal{D}^j_{[\mathbf{\Phi}(P\otimes I_n),\mathbf{\Phi}']}\left(\begin{bmatrix} \alpha_1 p_{\mathbf{s}_{t-j-1},1}\mathbf{v}'_{t-l}\phi(t-j-1,l-j-1)'\phi'_1 \\ \vdots \\ \alpha_K p_{\mathbf{s}_{t-j-1},K}\mathbf{v}'_{t-l}\phi(t-j-1,l-j-1)'\phi'_K \end{bmatrix}\right) \\
&\quad = \mathcal{D}^j_{[\mathbf{\Phi}(P\otimes I_n),\mathbf{\Phi}']}(\mathbf{A}\,\mathrm{diag}_K[P\xi_{t-j-1}\mathbf{v}'_{t-l}\phi(t-j-1,l-j-1)']\mathbf{\Phi}'J') \\
&\quad = \mathcal{D}^j_{[\mathbf{\Phi}(P\otimes I_n),\mathbf{\Phi}']}(\mathbf{A}\mathcal{D}_{[P,\mathbf{\Phi}']}(\xi_{t-j-1}\mathbf{v}'_{t-l}\phi(t-j-1,l-j-1)')).
\end{aligned}$$

Then (36) follows from (25) and the fact that $E(\xi_t \mathbf{v}'_t) = \mathrm{diag}_K(\pi)\mathbf{A}'J'$. ■

Proof of Theorem 2. The $\mathcal{L}_2$-boundedness of $\mathbf{x}_t$ is implied by the stated conditions. As the covariance-stationarity of $\mathbf{x}_t$ is given by (15)–(17) that are clearly independent of the time index t, it suffices to prove these expressions under the stated conditions. First, we note the fact that

$$E(\xi_t \otimes \mathbf{x}_t) = \sum_{j=1}^{\infty} E[\xi_t \otimes (\phi(t,j)\mathbf{v}_{t-j})]$$

and $\mathbf{x}_t = J(\xi_t \otimes \mathbf{x}_t)$. Then, (15) is shown to be true by (32) and Lemma 3 (note that $I_{nK} - \mathbf{\Phi}(P \otimes I_n)$ is invertible when $\|\mathbf{\Phi}(P \otimes I_n)\| < 1$). Further, because

$$\xi_t \otimes (\mathbf{x}_t\mathbf{x}'_t) = \sum_{j=0}^{\infty} \xi_t \otimes \psi_t(j,j) + \sum_{j=0}^{\infty}\sum_{l=j+1}^{\infty} (\xi_t \otimes \psi_t(l,j) + \xi_t \otimes \psi_t(l,j)')$$

and $\mathbf{x}_t\mathbf{x}'_t = J[\xi_t \otimes (\mathbf{x}_t\mathbf{x}'_t)]$, (16) follows from (34)–(36). The convergence of (16) is secured by Theorem 1. To show (17), we write

$$\mathbf{x}_{t+\tau}\mathbf{x}'_t = \phi(t+\tau,\tau)\mathbf{x}_t\mathbf{x}'_t + \sum_{j=0}^{\tau-1} \phi(t+\tau,j)\mathbf{v}_{t+\tau-j}\mathbf{x}'_t. \tag{38}$$

It follows from (24) that

$$E[\phi(t+\tau,\tau)\mathbf{x}_t\mathbf{x}'_t] = J[\mathbf{\Phi}(P \otimes I_n)]^{\tau}E[\xi_t \otimes (\mathbf{x}_t\mathbf{x}'_t)],$$

https://doi.org/10.1017/S026646660016102X Published online by Cambridge University Press

which results in the first term in (17). Using (24) again, we find

$$
\begin{aligned}
&E[\phi(t+\tau,j)\mathbf{v}_{t+\tau-j}\mathbf{x}_t'] \\
&\quad = J[\mathbf{\Phi}(P\otimes I_n)]^j E[\xi_{t+\tau-j}\otimes(\mathbf{v}_{t+\tau-j}\mathbf{x}_t')] \\
&\quad = J[\mathbf{\Phi}(P\otimes I_n)]^j \mathbf{A}P^{\tau-j}E[\xi_t\mathbf{x}_t'] \\
&\quad = J[\mathbf{\Phi}(P\otimes I_n)]^j \mathbf{A}P^{\tau-j}\sum_{a=0}^{\infty} E[\xi_t\mathbf{v}_{t-a}'\phi(t,a)'] \\
&\quad = J[\mathbf{\Phi}(P\otimes I_n)]^j \mathbf{A}P^{\tau-j}\sum_{a=0}^{\infty} \operatorname{diag}_K\{[\mathbf{\Phi}(P\otimes I_n)]^a\mathbf{A}\pi\}'J' \\
&\quad = J[\mathbf{\Phi}(P\otimes I_n)]^j \mathbf{A}P^{\tau-j}\operatorname{diag}_K\{\bar{\mathbf{A}}\}'J',
\end{aligned}
$$

where the fourth equality follows from (33) and the second equality follows from (23) and the fact that $\xi_t\otimes\mathbf{v}_t$ is a vector and $(\xi_t\otimes\mathbf{v}_t)\mathbf{x}_t' = \xi_t\otimes(\mathbf{v}_t\mathbf{x}_t')$. This and (38) result in the second term in (17) and complete the proof. ■

Proof of Theorem 3. Noticing that

$$
\mu = J[\mathbf{\Phi}(P\otimes I_n)]^\tau\bar{\mathbf{A}} + \sum_{j=0}^{\tau-1}[\mathbf{\Phi}(P\otimes I_n)]^j\mathbf{A}\pi,
$$

we find

$$
\begin{aligned}
\mathbf{\Gamma}_\tau &= J[\mathbf{\Phi}(P\otimes I_n)]^\tau(\bar{\mathbf{\Gamma}} - \bar{\mathbf{A}}\mu') \\
&\quad + J\sum_{j=0}^{\tau-1}[\mathbf{\Phi}(P\otimes I_n)]^j\mathbf{A}(P^{\tau-j}\operatorname{diag}_K(\bar{\mathbf{A}})'J' - \pi\mu'). \qquad \textbf{(39)}
\end{aligned}
$$

Under the stated conditions, it is evident that the first term of the preceding expression is of the magnitude $O(\|\mathbf{\Phi}(P\otimes I_n)\|^\tau)$. Further, we write

$$
\begin{aligned}
(J\operatorname{diag}_K(\bar{\mathbf{A}})P'^{\tau-j} - \mu\pi') &= J(\operatorname{diag}_K(\bar{\mathbf{A}})P'^{\tau-j} - \bar{\mathbf{A}}\pi') \\
&= J\begin{bmatrix} \bar{a}_1(q_{11}-\pi_1) & \dots & \bar{a}_1(q_{1K}-\pi_K) \\ \vdots & & \vdots \\ \bar{a}_K(q_{K1}-\pi_1) & \dots & \bar{a}_1(q_{KK}-\pi_K) \end{bmatrix} \\
&= JB(\tau-j),
\end{aligned}
$$

where q_{ik} is the (i,k)th element of $P'^{\tau-j}$ and $\bar{a}_i$ is the ith subvector of $\bar{\mathbf{A}}$. Under Assumption 3, $|q_{ik}-\pi_k| \le c\theta_1^{\tau-j}$ for all i and k, where the constants $\theta_1 \in (0,1)$ and $c > 0$ (see Doob, 1953). Therefore, the second term in (39) has a magnitude no greater than

https://doi.org/10.1017/S026646660016102X Published online by Cambridge University Press

$$\|J\| \sum_{j=0}^{\tau-1} \|\mathbf{\Phi}(P \otimes I_n)\|^j \|\mathbf{A}\| \|B(\tau - j)'\| \|J'\|$$

$$\leq c\|J\|^2 \|\mathbf{A}\| \|\bar{\mathbf{A}}\| \sum_{j=0}^{\tau-1} (\max\{\|\mathbf{\Phi}(P \otimes I_n)\|, \theta_1\})^\tau$$

$$= c\|J\|^2 \|\mathbf{A}\| \|\bar{\mathbf{A}}\| (\tau\delta^\tau),$$

where $\delta = \max\{\|\mathbf{\Phi}(P \otimes I_n)\|, \theta_1\} < 1$, concluding that $\|\mathbf{\Gamma}_\tau\| = O(\tau\delta^\tau)$ as $\tau \to \infty$. ■

Proof of Theorem 4. Let $\mathcal{W}_t$ be the space spanned by $\{\mathbf{w}_s : s \leq t\}$, which contains all finite linear combinations of $\{\mathbf{w}_s : s \leq t\}$ and their mean square limits. Let the linear projection of $\mathbf{x}_t$ on $\mathcal{W}_t$ be $\eta_t = \sum_{j=0}^{\infty} C_j \mathbf{w}_{t-j}$. The fact that $\mu_t = (\mathbf{x}_t - \eta_t) \perp \mathcal{W}_t$ implies that $\mu_t \perp \mathbf{w}_s$ for all $s \leq t$. But $E(\mu_t \mathbf{w}_s') = E[\mu_t E(\mathbf{w}_s' | \mathcal{F}_t)] = 0$ for all $s > t$, concluding that $\{\mu_t\}$ and $\{\mathbf{w}_t\}$ are uncorrelated. Further, the projection coefficient matrix is determined by $E(\mathbf{x}_t \mathbf{w}_{t-j}') = C_j E(\mathbf{w}_{t-j} \mathbf{w}_{t-j}')$ or $E(\mathbf{x}_{t+\tau} \mathbf{w}_t') \Omega^{-1} = C_\tau$ where $\Omega = E(\mathbf{w}_t \mathbf{w}_t')$ and the inverse is understood to be the generalized inverse. Note that $C_0 = E(\mathbf{x}_t \mathbf{w}_t') \Omega^{-1} = E[(\mathbf{x}_t - \mathbf{x}_{t|t-1}) \mathbf{w}_t'] \Omega^{-1} = I_n$, where $\mathbf{x}_{t|s} = E(\mathbf{x}_t | \mathcal{F}_s)$. Moreover, for

$$E(\mu_t | \mathcal{F}_{t-1}) = \mathbf{x}_{t|t-1} - \sum_{j=1}^{\infty} C_j \mathbf{w}_{t-j} = (\mathbf{x}_t - \mathbf{w}_t) - \sum_{j=1}^{\infty} C_j \mathbf{w}_{t-j} = \mathbf{x}_t - \sum_{j=0}^{\infty} C_j \mathbf{w}_{t-j}$$

$$= \mu_t,$$

the error of predicting μ_t is zero, and hence μ_t is purely deterministic. We therefore conclude that $\mu_t = E(\mu_t) = \mu$. To show the decay rate of C_τ, we only need to show that $E(\mathbf{x}_{t+\tau} \mathbf{w}_t')$ decays at the rate $O(\tau\delta^\tau)$. Denote $\mathbf{x}_{t|s} = E(\mathbf{x}_t | \mathcal{F}_s)$ and note that

$$E(\mathbf{x}_{t+\tau} \mathbf{w}_t') = E(\mathbf{x}_{t+\tau} \mathbf{x}_t') - E(\mathbf{x}_{t+\tau} \mathbf{x}_{t|t-1}')$$

$$= \mathbf{\Gamma}_\tau + \mu\mu' - E(\mathbf{x}_{t+\tau|t-1} \mathbf{x}_{t|t-1}')$$

$$= \mathbf{\Gamma}_\tau - E[(\mathbf{x}_{t+\tau|t-1} - \mu)(\mathbf{x}_{t|t-1} - \mu)'] . \tag{40}$$

We also note that

$$(\mathbf{x}_{t+\tau|t-1} - \mu) = \sum_{j=0}^{\tau} J[\mathbf{\Phi}(P \otimes I_n)]^j \mathbf{A} P(P^{\tau-j} \xi_{t-1} - \pi)$$

$$+ J[\mathbf{\Phi}(P \otimes I_n)]^{\tau+1} [(\xi_{t-1} \otimes \mathbf{x}_{t-1}) - \bar{\mathbf{A}}]$$

and

$$(\mathbf{x}_{t|t-1} - \mu) = J\mathbf{A}P(\xi_{t-1} - \pi) + J[\mathbf{\Phi}(P \otimes I_n)][(\xi_{t-1} \otimes \mathbf{x}_{t-1}) - \bar{\mathbf{A}}].$$

https://doi.org/10.1017/S026646660016102X Published online by Cambridge University Press

It can be further shown that

$$
\begin{aligned}
&E[(\mathbf{x}_{t+\tau|t-1} - \mu)(\mathbf{x}_{t|t-1} - \mu)'] \\
&\quad = \sum_{j=0}^{\tau} J[\mathbf{\Phi}(P \otimes I_n)]^j \mathbf{A} P \\
&\qquad \times \{[P^{\tau-j}\,\mathrm{diag}_K(\pi) - \pi\pi']P'\mathbf{A}'J' + [P^{\tau-j}\,\mathrm{diag}_K(\bar{\mathbf{A}}) - \pi\bar{\mathbf{A}}'] \\
&\qquad\quad \times [\mathbf{\Phi}(P \otimes I_n)]'J'\} + J[\mathbf{\Phi}(P \otimes I_n)]^{\tau+1} \\
&\qquad \times \{[\mathrm{diag}_K(\bar{\mathbf{A}}) - \mathbf{A}\pi']P'\mathbf{A}'J' + [\mathrm{diag}_K(\bar{\mathbf{\Gamma}}) - \bar{\mathbf{A}}\bar{\mathbf{A}}'][\mathbf{\Phi}(P \otimes I_n)]'J'\}.
\end{aligned}
$$

From the same argument as used in the proof of Theorem 3, it follows that both

$$[P^{\tau-j}\,\mathrm{diag}_K(\pi) - \pi\pi'] \quad \text{and} \quad [P^{\tau-j}\,\mathrm{diag}_K(\bar{\mathbf{A}}) - \pi\bar{\mathbf{A}}']$$

are bounded by $c\theta_1^{\tau-j}$, where the constants $\theta_1 \in (0,1)$ and $c > 0$. Therefore, the term $E[(\mathbf{x}_{t+\tau|t-1} - \mu)(\mathbf{x}_{t|t-1} - \mu)']$ is of the order $O(\tau\delta^\tau)$, where $\delta = \max\{\|\mathbf{\Phi}(P \otimes I_n)\|, \theta_1\} < 1$. The proof is completed by noting that $\mathbf{\Gamma}_\tau$ in (40) is also of the order $O(\tau\delta^\tau)$. ■

REFERENCES

Davidson, J. (1994) *Stochastic Limit Theory*, pp. 294–298. New York: Oxford University Press.

Doob, J.L. (1953) *Stochastic Processes*, pp. 173–174. New York: Wiley.

Engel, C. & J.D. Hamilton (1990) Long swings in the dollar: Are they in the data and do markets know it? *American Economic Review* 80, 689–713.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1008.

Evans, M.D.D. & K.K. Lewis (1995) Do expected shifts in inflation affect estimates of the long-run Fisher relation? *Journal of Finance* 50 (1), 225–253.

Granger, C.W.J. (1980) Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics* 14, 227–238.

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

Hamilton, J.D. (1990) Analysis of time series subject to changes in regime. *Journal of Econometrics* 45, 39–70.

Hamilton, J.D. (1996) Specification testing in Markov-switching time series models. *Journal of Econometrics* 70, 127–157.

Hansen, B.E. (1992) The likelihood ratio test under non-standard conditions: Testing the Markov switching model of GNP. *Journal of Applied Econometrics* 7, S61–S82.

Hosking, J.R.M. (1981) Fractional differencing. *Biometrika* 68, 165–176.

Kim, C.-J. (1994) Dynamic linear models with Markov-switching. *Journal of Econometrics* 60, 1–22.

Lam, P.-s. (1990) The Hamilton model with a general autoregressive component: Estimation and comparison with other models of economic time series. *Journal of Monetary Economics* 26, 409–432.

https://doi.org/10.1017/S026646660016102X Published online by Cambridge University Press