Econometric Theory, **28**, 2012, 1–41.
doi:10.1017/S0266466611000119

NULL RECURRENT UNIT ROOT PROCESSES

TERJE MYKLEBUST
Sogn og Fjordane University College

HANS ARNFINN KARLSEN AND DAG TJØSTHEIM
University of Bergen

The classical nonstationary autoregressive models are both linear and Markov. They include unit root and cointegration models. A possible nonlinear extension is to relax the linearity and at the same time keep general properties such as nonstationarity and the Markov property. A null recurrent Markov chain is nonstationary, and β-null recurrence is of vital importance for statistical inference in nonstationary Markov models, such as, e.g., in nonparametric estimation in nonlinear cointegration within the Markov models. The standard random walk is an example of a null recurrent Markov chain.

In this paper we suggest that the concept of null recurrence is an appropriate nonlinear generalization of the linear unit root concept and as such it may be a starting point for a nonlinear cointegration concept within the Markov framework. In fact, we establish the link between null recurrent processes and autoregressive unit root models. It turns out that null recurrence is closely related to the location of the roots of the characteristic polynomial of the state space matrix and the associated eigenvectors. Roughly speaking the process is β-null recurrent if one root is on the unit circle, null recurrent if two distinct roots are on the unit circle, whereas the others are inside the unit circle. It is transient if there are more than two roots on the unit circle. These results are closely connected to the random walk being null recurrent in one and two dimensions but transient in three dimensions. We also give an example of a process that by appropriate adjustments can be made β-null recurrent for any $\beta \in (0, 1)$ and can also be made null recurrent without being β-null recurrent.

1. INTRODUCTION

The concept of a unit root process is central in modern econometrics. In its simplest form the unit root property refers to a scalar autoregressive moving average (ARMA) process whose characteristic polynomial has a root at one. Such a process is often termed an I(1) process, whereas processes with a double root are said to be $I(2)$, and so forth. There are a number of alternative definitions of $I(1)$ processes, and classic unit root processes could be seen as being based on partial

We are very grateful to an anonymous referee for two extremely careful and detailed reviews of earlier versions of the paper. These reports have led to very substantial improvements of the paper. We also thank the chief editor Peter Phillips for a number of comments that have improved our presentation by putting our results in better perspective. Address correspondence to Terje Myklebust, Sogn og Fjordane University College, Ali, Pb 133, NO-6851 Sogndal, Norway; e-mail: terjemy@hisf.no.

© Cambridge University Press 2011

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

sums of weakly dependent (sometimes, but not always, linear process) variables as considered originally in Phillips (1987). Many of these alternatives are mentioned in a recent paper by Davidson (2009).

Most of these concepts have been defined and analyzed within a linear framework. Weak dependence in the component innovations is by no means a linear concept, but the nonstationarity is linear in the classic unit root world. There is a need for time series models that are both nonlinear and nonstationary. This has led to a search for nonlinear extensions of unit root processes and of the linear cointegration concept. A number of approaches have been tried. Some recent references include Aparicio, Escribano, and Sipols (2006), Bec, Guay, and Guerre (2008), Escribano (2004), Escribano, Sipols, and Aparicio (2006), Juhl and Xiao (2005a, 2005b), Saikkonen (2005), and Saikkonen and Choi (2004). It is not easy to establish a statistical theory of estimation and testing for these extensions. Park and Phillips (1999, 2001) have studied inference in parametric models using local time of Brownian motion. Wang and Phillips (2009) have further developed the use of local time of Brownian motion to obtain results in nonparametric estimation, whereas Karlsen and Tjøstheim (2001) and Karlsen, Myklebust, and Tjøstheim (2007) have suggested using Markov theory with emphasis on nonparametric estimation. There have been several recent papers using the Markov approach, sometimes combined with martingale arguments (Chen, Gao, and Li, 2009; Gao, King, Lu, and Tjøstheim, 2009; Saikkonen, 2007; Sancetta, 2009; Schienle, 2007).

There is a need to establish a link between these developments and the theory for classic unit root processes. Our attempted link goes through the Markov concept null recurrence (or β-null recurrence), and the main purpose of this paper is to explore this link to some extent. More precisely, we discuss whether null recurrence is analogous to a unit root structure for the class of autoregressive processes, from which we suggest null recurrence as a possible generalization of the unit root concept. However, we have to emphasize that null recurrence does not supplant unit root in all situations. Clearly, the Markov framework is a restriction in itself, and it excludes many classic unit root processes mentioned in the first paragraph of this section, such as, e.g., the fractional processes. However, the Markov framework is a well-known powerful tool. And, after all, many important processes are Markov or can be rewritten as Markov processes, such as, e.g., the class of ARMA processes. Finally, as demonstrated in Karlsen et al. (2007) and Karlsen, Myklebust, and Tjøstheim (2010), it is possible to develop a theory of inference for nonlinear time series regression models with a nonlinear transformation of a Markov input process and with stationary innovations.

Kallianpur and Robbins (1954) already proved that the simplest unit root process, the random walk, is a recurrent Markov chain under quite general circumstances. In fact they proved that it is β-null recurrent with $\beta = \frac{1}{2}$, where the parameter β regulates the expected number of visits to a so-called small set C. Essentially, a Markov chain $\{X_t\}$ is said to be β-null recurrent if

$$\mathrm{E}_\lambda \left[\sum_{t=0}^{N} 1_C(X_t)\right] = N^\beta L(N)(1+o(1)),$$

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

where $0 < \beta < 1$ and $L(N)$ is a slowly varying function (see Section 2 for a more precise definition). The concept of β-null recurrence is essential in constructing the theory of statistical inference presented in Karlsen and Tjøstheim (2001). It is also a useful tool in nonlinear cointegration (cf. Karlsen et al., 2007) and in unit root testing (cf. Gao et al., 2009). One would therefore like to establish that the recurrence property (and in particular β-null recurrence) holds for linear unit root Markov processes. Such a framework is needed also in testing a linear cointegration model against a nonlinear one, because the distribution of the relevant test statistic under the null hypothesis of linearity is required. Actually, the testing theory in Gao et al. (2009) is restricted to random walk type models because of the absence of such a framework.

Our chief objective in this paper, then, is to classify the class of autoregressive unit root processes in terms of recurrent and β-null recurrent Markov chains. To our knowledge no such classification exists in the literature, and we are able to show in Theorem 4 in Section 3 that a large class of unit root processes are in fact β-null recurrent. We are also able to say something in the case where there is more than one unit root.

2. SOME MARKOV THEORY

Before we derive our main result, Theorem 4, we need to establish some notation and theory for Markov chains, in particular for stability concepts, such as positive recurrence, null recurrence, and β-null recurrence.

Let $\{X_t : t \geq 0\}$ be a Markov chain on the state space $(E, \mathcal{E})$ with transition probability kernel P. We assume that $\mathcal{E}$ is a countably generated σ-algebra. In this paper we usually have $E = R^n$ and $\mathcal{E} = \mathcal{B}^n$ the Borel σ-algebra on R^n. We write P_λ and E_λ for the probability measure and the corresponding expectation with λ being the initial measure for X_0. When the initial measure is δ_x, the point measure concentrated at x, we will write P_x and E_x. In particular if the initial measure is the point measure concentrated at zero, we write P_0 and E_0. Then

$$\mathrm{P}_x(X_t \in A) = P^t(x, A), \qquad t \geq 0, \qquad A \in \mathcal{E}, \qquad x \in E,$$

with $P(x, A) = P^1(x, A)$, and $P^0 = I$, the identity kernel.

For a discrete Markov chain irreducibility means that it is possible to go to any state from any state, even if not in one time step. Unfortunately, this concept is not so easily defined in general. Let φ be a σ-finite positive measure defined on $\mathcal{E}$, e.g., the Lebesgue measure. The Markov chain $\{X_t\}$ is said to be φ-irreducible if for $A \in \mathcal{E}$,

$$\sum_{t=1}^{\infty} P^t(x, A) > 0 \quad \text{for all } x \in E \text{ whenever } \varphi(A) > 0.$$

We also say that $\{X_t\}$ is irreducible if it is φ-irreducible for some φ, and such a φ is an irreducibility measure for $\{X_t\}$. If $\{X_t\}$ is φ-irreducible, there exists a

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

maximal irreducibility measure ψ (see Tweedie, 1976, p. 741; Nummelin, 1984, p. 13). That is, ψ is a maximal irreducibility measure if for $A \in \mathcal{E}$,

$$\psi(A) = 0 \quad \text{implies} \sum_{t=1}^{\infty} P^t(x, A) = 0 \qquad \psi - \text{a.e.}$$

The maximal measure ψ is unique up to equivalence of measures.

Example 1
Let $\{X_t\}$ be the random walk on the integers and let δ_Z be the counting measure; i.e., for a set A, δ_Z is the number of integers contained in A. Then $\{X_t\}$ is δ_Z-irreducible, and in fact δ_Z is the maximal irreducibility measure. To emphasize the difference between an irreducibility measure and the maximal irreducibility measure, note that any point measure δ_x where $x \in \mathcal{Z}$ is an irreducibility measure for $\{X_t\}$.

The Lebesgue measure is a maximal irreducibility measure for many familiar chains, such as the AR(1) process with Gaussian noise (see Chan and Tong, 1985).

For a measure λ and a measurable function f we define

$$\lambda f = \lambda(f) = \int f(x)\lambda(dx).$$

Let

$$\mathcal{E}^+ = \{A \in \mathcal{E} : \psi(A) > 0\}.$$

If $\mathcal{E} = \mathcal{B}^n$, $\mathcal{B}^{n+}$ is defined similarly. For a nonnegative measurable function f, we use the notation $f \in \mathcal{E}^+$ if $\psi(f) > 0$. Moreover, the abbreviation i.o. is used for infinitely often.

DEFINITION 1. *Let $\{X_t\}$ be an irreducible Markov process on the state space $(E, \mathcal{E})$. We say that $\{X_t\}$ is recurrent if*

$$\mathrm{P}_x(X_t \in A \quad \text{i.o.}) = 1, \qquad \psi - \text{a.e.}, \quad \text{for all } A \in \mathcal{E}^+ \tag{2.1}$$

and transient (or dissipative) otherwise. The process $\{X_t\}$ is said to be Harris recurrent if (2.1) holds for all $x \in E$.

It is well known that if $\{X_t\}$ is recurrent, then there exists a σ-finite invariant measure π (see Nummelin, 1984, Cor. 5.2, p. 74); i.e.,

$$\pi(A) = \int \pi(dx) P(x, A) \quad \text{for all } A \in \mathcal{E}.$$

The invariant measure π is unique up to normalization, and it serves as a maximal irreducibility measure for the chain; i.e., π is equivalent to any maximal irreducibility measure ψ.

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

The first hitting time of a set A (it is a stopping time) is defined by

$$S_A = \inf\{t \geq 1 : X_t \in A\}, \qquad A \in \mathcal{E}.$$

Positive recurrence essentially means that the expected time until a process returns to a set A is finite. This is defined more precisely in Definition 2.

DEFINITION 2. *Suppose $\{X_t\}$ is recurrent. Then $\{X_t\}$ is said to be positive recurrent if*

$$\int_A \pi(dx)\mathrm{E}_x S_A < \infty \quad \textit{for all } A \in \mathcal{E}^+$$

and null recurrent otherwise.

For example, the AR(1) process $X_t = \phi X_{t-1} + \varepsilon_t$ where, for simplicity, the innovations are independent and identically distributed (i.i.d.) zero mean normally distributed variables, is positive recurrent for $|\phi| < 1$, transient for $|\phi| > 1$, and null recurrent for $|\phi| = 1$. In particular positive recurrence and transience are often easily proved by the so-called drift criteria; cf. Meyn and Tweedie (1993). The drift criteria are typically more difficult to apply in the null recurrent situations.

If the chain is positive recurrent, the invariant measure π is finite and is made unique by normalizing it to a probability measure. We refer to it as the stationary distribution of the Markov chain. Moreover, an aperiodic and positive Harris recurrent Markov chain $\{X_t\}$ is called ergodic (cf. Nummelin, 1984, p. 114); i.e.,

$$\lim_{t\to\infty} r^t ||P^t(x,\cdot) - \pi|| = 0 \qquad \text{for all } x \in E, \tag{2.2}$$

where $r \geq 1$ and $||\cdot||$ is the total variation norm. The process is called geometric ergodic if (2.2) holds for an $r > 1$ (see Nummelin, 1984, p. 120).

We need the concept of a small set, which in some sense corresponds to a single state in a discrete state space model. A function $\eta \in \mathcal{E}^+$ is small if there exist an integer $m \geq 1$, a constant $\gamma > 0$, and a nontrivial measure λ such that for any $A \in \mathcal{E}$ and any $x \in E$,

$$P^m(x, A) \geq \gamma\, \eta(x)\lambda(A). \tag{2.3}$$

We say that C is a small set if $1_C(x)$ is a small function. If λ satisfies (2.3) for some $\eta \in \mathcal{E}^+$, $\gamma > 0$, and $m \geq 1$, then λ is said to be a small measure. Condition (2.3) is often referred to as the minorization condition.

It can be shown (see Nummelin, 1984, Thm. 2.1) that if $\{X_t\}$ is irreducible, there exist a small function s, a probability measure ν, and an integer $m_0 \geq 1$ such that

$$P^{m_0}(x, A) \geq s(x)\nu(A), \qquad x \in E, \qquad A \in \mathcal{E}. \tag{2.4}$$

If the AR(1) process $X_t = \phi X_{t-1} + \varepsilon_t$ is again taken as an example, and if f is the density function of ε_t, then with $m_0 = 1$, it is easily seen that in (2.4) we may take $s(x) = a1_C(x)$, where 1_C is the indicator function of a compact set C, $a = \int \rho_0(y)dy$ with $\rho_0(y) = \inf_{x\in C} f(y - \phi x)$ and $\nu(dy) = a^{-1}\rho_0(y)dy$.

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

The small sets serve as "test" sets for recurrence and transience. In many situations, or we might say for the most common processes, subsets of compact sets are small if they have positive Lebesgue measure (see, e.g., Chan and Tong, 1985). Moreover, any finite union of small sets is a small set (see Nummelin, 1984, Cor. 2.1, p. 24).

The well-known result (see Nummelin, 1984, p. 77)

$$\int \pi(dx)\mathrm{E}_x S_A = \pi(E) \qquad \text{for all } A \in \mathcal{E}^+$$

gives that the invariant measure π is infinite in the case of null recurrence. But still, $\pi(s)$ is finite for any small function s, and thus $\pi(C)$ is finite for any small set C (see Nummelin, 1984, Prop. 5.6, p. 72). We write π_s for the invariant measure that is normalized in such a way that

$$\pi_s(s) = 1, \tag{2.5}$$

where s satisfies (2.4).

The two following results will be useful in proving our main theorem.

THEOREM 1. *Let $\{X_t\}$ be an irreducible Markov process.*

(i) Then $\{X_t\}$ is recurrent if and only if for any $A \in \mathcal{E}^+$

$$G(x, A) \stackrel{def}{=} \sum_{t=0}^{\infty} P^t(x, A) = \infty \qquad \textit{for all } x \in E,$$

which is equivalent to the following condition: there exists a small function s, such that

$$Gs(x) = \int G(x, dy)s(y) = \infty \qquad \textit{for some } x \in E.$$

(ii) The process $\{X_t\}$ is recurrent if and only if for some small set C,

$$\mathrm{P}_x(S_C < \infty) = 1, \qquad \psi - \text{a.e. } x \in C, \tag{2.6}$$

and $\{X_t\}$ is Harris recurrent if and only if (2.6) is true for all $x \in C$.

This result is an excerpt from Theorem 3.7 in Nummelin (1984, p. 43). Our result contains parts (ii), (iii), (vi), and (vii) of his Theorem 3.7. Note that Nummelin uses the notation $U_C 1(x) = \mathrm{P}_x(S_C < \infty)$ (cf. eq. (3.12), p. 34 in his book), and that his Definition 3.2 (pp. 27–28) is needed to read Theorem 3.7 in his book.

The next lemma says that if $P^t(x, A)$ goes to zero sufficiently fast, the chain is null recurrent.

LEMMA 1. *Let $\{X_t\}$ be an irreducible recurrent Markov process. The process $\{X_t\}$ is null recurrent if there exist an integer $k \geq 2$, a ψ-positive set $A \in \mathcal{E}^+$, and a state $x \in E$, such that*

$$\sum_{t=0}^{\infty} [P^t(x, A)]^k < \infty.$$

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

Proof. Define $\{Z_t\} = \{(X_t^{(1)}, \ldots, X_t^{(k)})\}$, $t \geq 0$, where $\{X_t^{(1)}\}, \ldots, \{X_t^{(k)}\}$ are independent copies of $\{X_t\}$. Because of the independence, the t-step transition kernel $\underline{P}^t$ of $\{Z_t\}$ is the product of the t-step transition kernels for the marginal chains, i.e., $\underline{P}^t = P^t \times \cdots \times P^t$. Moreover, the invariant measure for the coupled chain, $\{Z_t\}$, is $\underline{\pi} = \pi \times \cdots \times \pi$, where π is the invariant measure for $\{X_t\}$. Thus $\underline{\pi}$ is a finite measure if and only if π is a finite measure. By Nummelin (1984, Def. 5.3, p. 73, and Cor. 5.3, p. 77) the invariant measure is finite if and only if the chain is positive recurrent, which means that the coupled chain is positive recurrent if and only if $\{X_t\}$ is positive recurrent.

Now, let $\underline{x} = (x, \ldots, x)$ and $\underline{A} = A \times \cdots \times A$. If $\sum_{t=0}^{\infty} \underline{P}^t(\underline{x}, \underline{A}) = \sum_{t=0}^{\infty} \left[P^t(x, A)\right]^k$ is finite, then according to Theorem 1, $\{Z_t\}$ is transient. But the coupled chain $\{Z_t\}$ is positive recurrent if the marginal chains are positive recurrent. Therefore $\{X_t\}$ is null recurrent. ■

Many of the asymptotic statistics for null recurrent Markov models require regularity of tail behavior of the first hitting time. In this context we have used the term β-null recurrence for this regularity. An operational definition is given in Definition 3, which follows.

For two sequences $\{a_n; n \geq 0\}$ and $\{b_n; n \geq 0\}$ we write $a_n \sim b_n$ if $a_n/b_n \to 1$ as $n \to \infty$. A positive function L defined on $[a, \infty)$, where $a \geq 0$, is a slowly varying function at infinity (Bingham, Goldie, and Teugels, 1987, p. 6) if

$$\lim_{x \uparrow \infty} \frac{L(\kappa x)}{L(x)} = 1 \quad \text{for all } \kappa > 0. \tag{2.7}$$

Logarithmic functions are typical examples of slowly varying functions.

DEFINITION 3. *An irreducible, null recurrent Markov chain $\{X_t\}$ is β-null recurrent if there exist a small function h, an initial measure λ, a constant $\beta \in (0, 1)$, and a slowly varying function L_s (depending on s, defined by (2.4)) such that*

$$\mathrm{E}_\lambda \left[\sum_{t=0}^{N} h(X_t) \right] \sim \frac{1}{\Gamma(1+\beta)} N^\beta L_s(N) \pi_s(h) \quad \text{as } N \to \infty. \tag{2.8}$$

Remark 1. First we note that the leading term of the right-hand side of (2.8) does not depend on the initial measure λ. Hence this must also be true for the left-hand side. At the same time the left-hand side of (2.8) is independent of s, and therefore this is also true for the right-hand side. But π_s depends upon s, and then necessarily L_s depends on s. The factorization $L_s/\Gamma(1+\beta)$ in this formula is somewhat arbitrary because any $\Gamma(1+\beta)$ could have been included in the slowly varying function L_s. However, it is convenient because $\sum_{t=0}^{N} h(X_t)/N^\beta L_s(N)$ $\pi_s(h)$ is asymptotically Mittag Lefler distributed (see Karlsen and Tjøstheim, 2001, Thm. 3.2). By solidarity properties of the chain the condition (2.8) eventually is true for all small functions.

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

The most important example is the smooth simple random walk with innovation variance σ^2. It is β-null recurrent with $\beta = 1/2$ (see Kallianpur and Robbins, 1954) and

$$\mathrm{E}_\lambda\left[\sum_{t=0}^{N} h(X_t)\right] = \frac{2}{\sqrt{2\pi}}\frac{1}{\sigma}\Big\{\int h(x)dx\Big\}N^{1/2}\big(1+o(1)\big)$$
$$= \frac{1}{\Gamma(1+\beta)}L_s(N)\pi_s(h)N^{1/2}\big(1+o(1)\big)$$

with

$$L_s(N) = \frac{1}{\sigma\sqrt{2}}\int s(x)dx, \qquad \pi_s(h) = \frac{\int h(x)dx}{\int s(x)dx}.$$

Here L_s is constant and π_s is proportional to the Lebesgue measure.

If h is the indicator of a small set C then the right-hand side of (2.8) is the expected number of visits to the set C during $N+1$ time steps. This is also pathways approximately true, i.e., (2.8) without expectation. Note that in the positive recurrent case the corresponding number of visits is a fraction of N by the ergodic theorem (2.2). This is quite different from the random walk case where this number is a constant times the square root of N.

It is possible to formulate (2.8) as a regularly varying tail condition on the distribution of a first hitting time, but for a chain with continuous state space this requires the split chain, which is not discussed here. Details of this approach can be found in Karlsen and Tjøstheim (2001).

DEFINITION 4. *For a stochastic chain* $\{X_t; t \geq 0\}$ *we define the* m*-step chain* $\{X_t^m; t \geq 0\}$ *by* $\{X_t^m = X_{tm}; t \geq 0\}$ *where* m *is some positive integer.*

The next result shows that the m-step chain and the original chain are equivalent with respect to β-null recurrence. This is an important property for handling β-null recurrence because, as for positive recurrence (cf. Tjøstheim, 1990), the m-step chain is often easier to work with.

LEMMA 2. *The original chain,* $\{X_t\}$*, is* β*-null recurrent if and only if there exists an* $m > 1$ *such that the* m*-step chain,* $\{X_t^m\}$*, is* β*-null recurrent. If there exists such an* m*, then in fact* $\{X_t^m\}$ *is* β*-null recurrent for all* m*.*

If L_s^m *denotes the slowly varying function for the* m*-step chain then*

$$L_s^m = m^{\beta-1}L_s\,. \tag{2.9}$$

If $m_0 = 1$ *(*m_0 *was defined by (2.4)), then the split chain of* $\{X_t\}$*, which is defined in Nummelin (1984, p. 60), is* β*-null recurrent.*

Proof. Let $G^{(N)} \stackrel{\text{def}}{=} \sum_{t=0}^{N-1} P^t$. We have

$$G^{(Nm-1)} = \sum_{t=0}^{N-1}\sum_{k=0}^{m-1} P^{tm+k} = \sum_{t=0}^{N-1} P^{tm}\sum_{k=0}^{m-1} P^k = G_m^{(N)}\Delta_m, \tag{2.10}$$

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

where $G_m^{(N)} \stackrel{\text{def}}{=} \sum_{t=0}^{N-1} P^{tm}$ and $\Delta_m \stackrel{\text{def}}{=} \sum_{k=0}^{m-1} P^k$. Note that $\Delta_m h$ is a small function if h is small (see Nummelin, 1984, Prop. 2.10, p. 22). This identity links the m-step chain to the original chain by (2.8), and so $\{X_t^m\}$ is β-null recurrent for all $m \geq 1$ if $\{X_t\}$ is.

To verify (2.9), assume that the original chain is β-null recurrent and let h be a small function. Then, because $\pi_s(\Delta_m h) = m\pi_s h$, we get

$$\begin{aligned}\lambda G_m^{(N)} \Delta_m h &= \lambda G^{(Nm+m-1)} h \\ &\sim \frac{\pi_s h}{\Gamma(1+\beta)} (Nm)^\beta L_s(Nm) \\ &\sim \frac{\pi_s(\Delta_m h)}{\Gamma(1+\beta)} N^\beta \{m^{\beta-1} L_s(Nm)\}. \end{aligned} \tag{2.11}$$

By (2.7) $L_s(Nm) \sim L_s(N)$, and (2.9) follows.

Now, assume that $\{X_t^m\}$ is β-null recurrent for some $m > 1$. Let J be an integer. We can write $J = [J/m]m - 1 + l$, where $[\cdot]$ is the integer function and where $1 \leq l \leq m$. Now as $J \to \infty$, by (2.10) and (2.11) we have that

$$\begin{aligned}\lambda G^{(J)} h &= \lambda G^{([J/m]m-1+l)} \\ &= \lambda G_m^{([J/m])} \Delta_m h + \sum_{k=[J/m]m}^{J} \lambda P^k h \\ &\sim \lambda G_m^{([J/m])} \Delta_m h \\ &\sim \frac{\pi_s(\Delta_m h)}{\Gamma(1+\beta)} ([J/m])^\beta \{m^{\beta-1} L_s([J/m])\} \\ &\sim \frac{\pi_s(\Delta_m h)}{\Gamma(1+\beta)} J^\beta m^{-1} L_s(J) \\ &\sim \frac{\pi_s h}{\Gamma(1+\beta)} J^\beta L_s(J). \end{aligned} \tag{2.12}$$

This shows that $\{X_t\}$ is β-null recurrent.

The last statement of the lemma is obvious because $C \times \{0, 1\}$ is a small set for the split chain (for a definition, see Nummelin, 1984, Ch. 4) when C is small for the original chain. ■

Let $\{X_t\}$ be an irreducible recurrent Markov chain and let C be a small set. If there are an $x \in E$, a constant c, and a $\beta \in (0, 1)$ such that

$$\mathrm{P}_x(X_t \in C) \sim c t^{-(1-\beta)}, \quad \text{then } \{X_t\} \text{ is } \beta\text{-null recurrent.} \tag{2.13}$$

Null recurrence follows by Lemma 1, and β-null recurrence follows from (2.8) with $\lambda = \delta_x$, the point measure at x, and the relationship

$$\sum_{t=0}^{N} t^{-(1-\beta)} \sim \beta^{-1} N^\beta \quad (0 < \beta < 1). \tag{2.14}$$

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

An important property of null recurrence is stated in the following theorem.

THEOREM 2. *Let $\{X_t\}$ be an irreducible β-null (null) recurrent Markov chain on the state space $(E, \mathcal{E})$. Suppose g is a measurable one-to-one transformation; then $\{Y_t = g(X_t)\}$ is an irreducible β-null (null) recurrent Markov chain.*

Proof. For any set $A \in \mathcal{E}$ define

$$\tilde{A} = \{y : y = g(x), \quad x \in A\}$$

and for any measure λ, on the state space $(E, \mathcal{E})$, define $\tilde{\lambda}$ such that

$$\tilde{\lambda}(\tilde{A}) = \lambda(A) \qquad \forall A \in \mathcal{E}.$$

It is clear that $\{Y_t\}$ is $\tilde{\lambda}$-irreducible if and only if $\{X_t\}$ is λ-irreducible. Moreover,

$$\mathrm{E}_{\tilde{\lambda}}\left[\sum_{t=0}^{N} 1_{\tilde{A}}(Y_t)\right] = \mathrm{E}_{\lambda}\left[\sum_{t=0}^{N} 1_A(X_t)\right],$$

which shows that $\{Y_t\}$ is β-null (null) recurrent if and only if $\{X_t\}$ is β-null (null) recurrent. ■

Theorem 2 shows that β-null recurrence is an inherent nonlinear concept invariant to one-to-one nonlinear transformations. This is in contrast to the $I(1)$ concept where lack of invariance may create difficulties for $I(1)$ as a nonlinear concept (cf. Granger and Hallman, 1991a, 1991b; Granger, 1995; Granger and Swanson, 1996).

We close this section by studying a class of chains that provides examples of β-null recurrence for any β-value in the interval $(0, 1)$ and also null recurrent chains that are not β-null recurrent. The example also demonstrates that null (and β-null) recurrence may depend on "local" properties of the transition probability kernel; i.e., null recurrence can be due to the transition probability $P(x, \cdot)$ for x in a small set.

Example 2

Let η_t, $t \geq 1$, be a sequence of i.i.d. variables taking positive values and let $\{X_t; t \geq 0\}$ be a process with state space $(R_+, \mathcal{B})$, where R_+ is the nonnegative real line. Moreover, let $\{X_t\}$ be defined recursively by

$$X_t = \begin{cases} X_{t-1} - 1, & \text{if } X_{t-1} > 1, \\ \eta_t, & \text{if } X_{t-1} \in [0, 1]. \end{cases} \tag{2.15}$$

Then $\{X_t\}$ is easily seen to be an aperiodic l_+-irreducible Markov chain where l_+ is the Lebesgue measure constrained to the nonnegative real line. Also, note that the set $[0, 1]$ is a proper atom. Moreover, we have that for $S_{[0,1]} = \inf\{t \geq 1 : X_t \in [0, 1]\}$,

$$\mathrm{P}_x(S_{[0,1]} = [x] + 1) = 1 \quad \text{for } x > 1, \tag{2.16}$$

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

where $[\cdot]$ is the integer function. Thus

$$\mathrm{P}_x(S_{[0,1]} > n) = \mathrm{P}([\eta_1] > n) \quad \text{for } x \in [0,1], \tag{2.17}$$

which means that $\{X_t\}$ is null recurrent if and only if

$$\mathrm{E}[\eta_1] = \infty \tag{2.18}$$

and β-null recurrent if and only if (see Karlsen and Tjøstheim, 2001, Thm. 3.1)

$$\mathrm{P}([\eta_1] > n) \sim n^{-\beta} L^{-1}(n), \qquad \beta \in (0,1), \tag{2.19}$$

where L is some slowly varying function.

It should be clear from (2.19) that we can construct examples of β-null recurrent chains for any $\beta \in (0,1)$ and also examples of null recurrent chains that are not β-null recurrent.

For a further discussion of the concept of β-null recurrence we refer to Karlsen and Tjøstheim (2001).

3. NULL RECURRENT VAR-PROCESSES

In this section we classify the vector autoregressive processes (VAR) in terms of recurrence properties.

It is well known that the random walk is null recurrent in one and two dimensions but transient in higher dimensions (see Spitzer, 1964, Sect. 8). However, to our knowledge, it is not known which VAR processes are null recurrent. In this section, we try to answer this question.

DEFINITION 5. *An n-dimensional* VAR(1) *process,* $\{X_t; t \geq 0\}$*, is given by*

$$X_t = \begin{cases} X_0, & \text{for } t = 0, \\ AX_{t-1} + b + \varepsilon_t, & \text{for } t > 0, \end{cases} \tag{3.1}$$

where X_t*,* b*, and* ε_t *have dimension* $n \times 1$ *and* A *is an* $n \times n$ *matrix. The noise process is i.i.d., independent of* X_0 *and* $\mathrm{E}[\varepsilon_1] = 0$*.*

The i.i.d. property of Definition 5 is a strong assumption. Independence is crucial in our approach, and identically distributed variables will ensure that the Markov process is time homogeneous. Note, though, that in some cases (e.g., an ARMA model) it is possible to redefine the state space to accommodate dependence in $\{\varepsilon_t\}$.

Remark 2. Note that the matrix A in Definition 5 is not assumed to be nonsingular. On the other hand if ε_1 has a density and a finite second moment, then the covariance matrix Σ_ε is positive definite.

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

Remark 3. A VAR(1) process is a Markov process with state space R^n, initial distribution λ equal to the distribution of X_0, and transition probability kernel

$$P(x, dy) = f_\varepsilon(y - Ax - b)dy,$$

when the noise ε has a density. Moreover, the VAR(1) process is a time-homogeneous Markov process because $\{\varepsilon_t\}$ is i.i.d.

It is known that A is similar to a Jordan matrix B, where the eigenvalues of A occur along the diagonal of B in nonincreasing order with respect to the absolute value, as one moves from left to right. Furthermore, B has the form

$$B = \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & J_k \end{bmatrix}, \qquad J_i = \begin{bmatrix} \gamma_i & 0 & \cdots & 0 & 0 \\ 1 & \gamma_i & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \gamma_i \end{bmatrix}, \qquad i = 1, \ldots, k;$$

i.e., $J_1, \ldots, J_k$, $(k \leq n)$, are elementary Jordan matrices with eigenvalues $\gamma_1, \ldots, \gamma_k$, respectively, and γ_i, $i = 1, \ldots, k$, are the eigenvalues of A ($\gamma_1, \ldots, \gamma_k$ are not necessarily distinct). If the eigenspace of A is n-dimensional, then $J_1, \ldots, J_n$ are scalars. This is surely satisfied if the eigenvalues of A are distinct. Suppose that $J_1, \ldots, J_l$, $(l \leq k)$, have their eigenvalues on the unit circle whereas $J_{l+1}, \ldots, J_k$ have their eigenvalues inside the unit circle. Then we can decompose B into two blocks such that

$$B = \begin{bmatrix} B_1 & 0 \\ 0 & B_2 \end{bmatrix}, \tag{3.2}$$

where (all) the diagonal element(s) of B_1 is (are) on the unit circle and the diagonal element(s) of B_2 is (are) inside the unit circle. In this paper, the symbols B, B_1, and B_2 will always refer to the representation (3.2).

In Theorem 4 we need some restrictions on the noise process $\{\varepsilon_t\}$ and on the constant vector b. Conditions C1–C3, which follow, ensure the existence of a local limit theorem for the densities. First, we define the characteristic function for a general stochastic vector ε of dimension $n \geq 1$,

$$g_\varepsilon(u) \stackrel{\text{def}}{=} \mathrm{E}[\exp(iu \cdot \varepsilon)], \tag{3.3}$$

where $u \cdot v = \sum_{i=1}^n u_i v_i$, for $u, v \in R^n$, is the euclidean inner product and $||u|| = (u \cdot u)^{1/2}$ is the euclidean norm.

We allow some of the components of the vector $\varepsilon_t' = [\varepsilon_{t,1}, \ldots, \varepsilon_{t,n}]$ to be zero, $t \geq 1$. Let $\varepsilon = \varepsilon_t$ with $t = 1$ and let $\tilde{\varepsilon}$ be the vector consisting of the nonzero

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

components of ε; note that $\tilde{\varepsilon}$ has dimension $\tilde{n} \leq n$. Then the vector $\tilde{\varepsilon}$ will satisfy C1–C3.

C1. The third moment is finite; i.e., $\mathrm{E}||\tilde{\varepsilon}||^3 < \infty$.

C2. The characteristic function is absolutely integrable; i.e., $\int |g_{\tilde{\varepsilon}}(u)|du < \infty$, where the integral is over $R^{\tilde{n}}$.

C3. The Cramér condition holds; i.e., $\sup_{||u||>a} |g_{\tilde{\varepsilon}}(u)| < 1$, for all $a > 0$, $u \in R^{\tilde{n}}$.

Remark 4. Condition C2 implies that $\tilde{\varepsilon}$ has a density, $f_{\tilde{\varepsilon}}$. Moreover, $f_{\tilde{\varepsilon}}$ is uniformly continuous and $f_{\tilde{\varepsilon}}(0) > 0$.

The following result is well known and will be used to prove Theorem 4.

THEOREM 3. *Assume that the* VAR(1) *process defined by Definition 5 is irreducible and that the eigenvalues of the matrix A are inside the unit circle, the variance exists, and Condition C2 holds. Then it is geometric ergodic; i.e., (2.2) holds for some* $r > 1$.

This result is a special case of Theorem 3 in Feigin and Tweedie (1985) and of Theorem 4.4 in Tjøstheim (1990). Note that geometric ergodicity is a shared property for any m-step chain $m \geq 1$ (see Nummelin, 1984, Thm. 6.14, p. 120). Therefore the assumptions in Feigin and Tweedie (1985) are satisfied by our Condition C2 through Lemma 3 and Lemma 8 in our Section 4. Note also that Theorem 3 in Feigin and Tweedie (1985) is formulated when the constant vector $b = 0$. However, this is not a problem because by standard transformation we can define $\tilde{X}_t = X_t - \mu$ where $\mu = (I - A)^{-1}b$, and we get the difference equation $\tilde{X}_t = A\tilde{X}_{t-1} + \varepsilon_t$.

In null recurrent models the matrix $(I - A)$ is in general not invertible, and we cannot apply the transformation mentioned previously. Therefore, the constant vector b has to be restricted to an appropriate subspace of R^n. For example, the one-dimensional random walk is transient if $b \neq 0$, i.e., $b \neq 0$ causes a drift. In the vector case (3.1) the process can be recurrent even if the vector b is nonzero, as long as it does not affect the integrated part of the process.

Now, let Q be the nonsingular $n \times n$ matrix such that $A = QBQ^{-1}$ where B is the matrix in (3.2) and let $q_1, \ldots, q_n$ be the column vectors of Q. A vector $b \in R^n$ can be written as

$$b = Q\alpha = \alpha_1 q_1 + \cdots + \alpha_n q_n, \quad \text{where } \alpha = [\alpha_1, \ldots, \alpha_n]' \in \mathcal{C}^n, \tag{3.4}$$

where $\mathcal{C}$ means the complex numbers. Recall that $\gamma_1, \ldots, \gamma_n$ are the n eigenvalues of the $n \times n$ matrix A; they might not all be distinct values. Define the real vector space

$$G_A = \{b \in R^n : b = \alpha_1 q_1 + \cdots + \alpha_n q_n, \text{ where } \alpha_i = 0 \text{ if } \gamma_i = 1\}. \tag{3.5}$$

So, e.g., if $\gamma_1 = \gamma_2 = 1$ are eigenvalues of the matrix A, then $b \in G_A$ means that $b = \alpha_3 q_3 + \cdots + \alpha_n q_n$, where $\alpha_3, \ldots, \alpha_n$ are complex numbers. Note that q_2 does

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

not need to be an eigenvector of A. This implies that $b \in G_A$ does not cause a drift in any of the components of the process $\{X_t\}$. So $b \in G_A$ is analogous to the well-known restriction on the constant vector in linear $I(1)$ analysis (cf. Hamilton, 1994, p. 581).

THEOREM 4. *Let l be the Lebesgue measure on R^n and let $\{X_t\}$ be an l-irreducible* VAR(1) *process of dimension n with state space $(R^n, \mathcal{B}^n)$ defined by (3.1). Assume that the noise process $\{\varepsilon_t\}$ satisfies C1–C3, where $\tilde{\varepsilon}$ is the nonzero component of ε_1. Furthermore, assume that the eigenvalues of A are on or inside the unit circle. Then we have that*

(i) $\{X_t\}$ *is β-null recurrent with $\beta = 1/2$ if $b \in G_A$ and A has exactly one eigenvalue on the unit circle. This eigenvalue is necessarily* 1 *or* -1 *because $\{X_t\}$ is assumed to be real.*

(ii) $\{X_t\}$ *is null recurrent, but not β-null recurrent for any $\beta \in (0, 1)$, if $b \in G_A$ and B_1, defined in (3.2), is diagonal and has dimension two; i.e., A has exactly two eigenvalues on the unit circle that might be equal, but the corresponding eigenspace is two-dimensional.*

(iii) $\{X_t\}$ *is transient if B_1 is a nondiagonal 2×2 matrix, i.e., B_1 is an elementary Jordan matrix (of dimension two) with characteristic value on the unit circle. In other words, A has two equal eigenvalues at* -1 *or at* 1 *whose eigenspace has dimension one.*

(iv) $\{X_t\}$ *is transient if the dimension of B_1 is higher than two; i.e., A has three or more eigenvalues on the unit circle.*

(v) $\{X_t\}$ *is transient if $b \notin G_A$.*

The proof is given in Section 4.

To provide some immediate intuition to the result, recall that a random walk is null recurrent in one and two dimensions but transient in higher dimensions. This is analogous with our result. Because the random walks in one and two dimensions have, respectively, one and two eigenvalues at unity, they correspond to parts (i) and (ii), respectively, of Theorem 4. The three-dimensional random walk has three eigenvalues at unity and therefore corresponds to part (iv) of Theorem 4. This is also similar to the recurrence properties of the continuous Brownian motion processes in the different dimensions.

Up to now we have only treated the first-order process, but higher order processes may be restated as first-order processes. The VAR(p) process of dimension $n \times 1$,

$$X_t = \alpha + \Phi_1 X_{t-1} + \cdots + \Phi_p X_{t-p} + \varepsilon_t, \tag{3.6}$$

where $\Phi_1, \ldots, \Phi_p$ are $n \times n$ matrices and α is an $n \times 1$ constant vector, can be written as a Markov process, i.e., in the form

$$Y_t = AY_{t-1} + b + \eta_t, \tag{3.7}$$

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

where the $pn \times pn$ matrix A is defined by

$$A = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{p-1} & \Phi_p \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix},$$

where I is the $n \times n$ identity matrix, 0 is the $n \times n$ matrix of zeros, $Y_t' = (X_t', \ldots, X_{t-p+1}')$, $b' = (\alpha', 0, \ldots, 0)$, and $\eta_t' = (\varepsilon_t', 0, \ldots, 0)$, $t \geq 0$, are vectors of dimension $pn \times 1$.

The following condition on the VAR(p) process (3.6) will ensure that the vectorization (3.7) is irreducible.

C4. The n-dimensional random noise vectors $\{\varepsilon_t; t \geq 1\}$ each have a strictly positive density with respect to the n-dimensional Lebesgue measure.

PROPOSITION 1. *Let $\{X_t\}$ be an n-dimensional* VAR(p) *process whose noise process satisfies Condition C4. Its state space Markov representation (3.7) is an irreducible Markov chain on the state space $(R^{pn}, \mathcal{B}^{pn})$.*

Proof. See Chan and Tong (1985). ■

We have the following corollaries to Theorem 4.

COROLLARY 1. *Let $\{X_t\}$ be a real valued* AR(p) *process satisfying C1–C3, so that the corresponding vectorization Y_t, of the form (3.7) with $b = 0$, is a p-dimensional* VAR(1) *process. Suppose the characteristic polynomial of $\{X_t\}$ has exactly one root on the unit circle and the other roots outside the unit circle. Then $\{Y_t\}$ is β-null recurrent, with $\beta = 1/2$.*

Proof. By Remark 4, Condition C4 is satisfied for the one-dimensional random noise vectors. Therefore the vectorization $\{Y_t\}$ on the state space $(R^p, \mathcal{B}^p)$ is irreducible by Proposition 1. Now the result follows by Theorem 4(i). ■

COROLLARY 2. *Let $\{X_t\}$ be a real valued* AR(p) *process satisfying (3.6) and assume that C1–C3 hold.*

(i) Let Y_t be the vectorization of X_t; i.e., Y_t is of the form (3.7). Assume that the constant vector $b \in G_A$. If the matrix A has two distinct eigenvalues on the unit circle and the other eigenvalues inside the unit circle, then $\{Y_t\}$ is null recurrent but not β-null recurrent.

(ii) The process $X_t = cX_{t-1} - X_{t-2} + \varepsilon_t$, with $c = -2$ or $c = 2$, written as a Markov process of the form (3.7) is transient. Note that $\{X_t\}$ is integrated of order two abbreviated $I(2)$ when $c = 2$, i.e., $(1 - L)^2 X_t = \varepsilon_t$ where L is the backward shift operator.

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

Proof. The vectorization $\{Y_t\}$ of $\{X_t\}$ is irreducible. It follows by the same arguments as in the proof of Corollary 1. Now (i) follows from Theorem 4(ii).

The matrix in part (ii) has two eigenvalues at 1 (or -1), and the eigenspace of this matrix is one-dimensional. So $B_1 = J$ is a two-dimensional elementary Jordan matrix with characteristic value 1 (or -1). Thus the result follows from Theorem 4(iii). ■

Transience means that the process visits some fixed compact set only a finite number of times, which indicates that any form of local estimation is impossible (cf. Karlsen and Tjøstheim, 2001). So it is interesting to note that the $I(2)$ process in part (ii) of Corollary 2 is transient. Nonparametric estimation is often interesting in a nonlinear situation, but this indicates that it might be difficult for a nonlinear counterpart to an $I(2)$ process.

The following example further illuminates part (iii) of Theorem 4. Suppose that the VAR(1) process

$$\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}, \tag{3.8}$$

where $b \neq 0$, satisfies Conditions C1–C3. Then it is transient by Theorem 4(iii). The first component of the process in (3.8) can be written as

$$(1-L)^2 X_{1,t} = \varepsilon_{1,t} - \varepsilon_{1,t-1} + b\varepsilon_{2,t},$$

where L is the backward shift operator; i.e., the first component of (3.8) is an $I(2)$ process. In fact the first component, $\{X_{1,t}\}$, can be rewritten as a Markov process without involving the second component directly. That is, $\{(X_{1,t}, X_{1,t-1}, \varepsilon_t)'\}$ is a Markov process,

$$\begin{bmatrix} X_{1,t} \\ X_{1,t-1} \\ \varepsilon_{1,t} \end{bmatrix} = \begin{bmatrix} 2 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{1,t-2} \\ \varepsilon_{1,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} + b\varepsilon_{2,t} \\ 0 \\ \varepsilon_{1,t} \end{bmatrix}. \tag{3.9}$$

The noise process of the two-step chain version of (3.9) has a positive definite covariance matrix, and so (3.9) is irreducible. And like the $I(2)$ processes in part (ii) of Corollary 2, (3.9) is transient by Theorem 4(iii).

Note that the process in (3.8) is not a cointegrated process because the first component is an $I(2)$ process rather than an $I(1)$ process. This is due to the fact that the matrix in (3.8) is essentially a Jordan matrix. Part (iii) of Theorem 4 handles the cases where there are multiple eigenvalues at 1 (or -1) whose eigenspace is degenerated; i.e., B_1 in (3.2) contains a Jordan matrix. The preceding example indicates that these processes contain processes that are $I(2)$ (or $I(d)$ $d \geq 2$) and therefore these processes are not cointegrated.

By Theorem 4(iv) a linear cointegrated process of dimension n with less than $n-2$ cointegrating relationships (more than two common trends) is transient.

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

This result is possibly a bit disappointing. It tells us that null recurrence does not work as a generalization of the linear unit root concept in all situations. That is, there are processes where each component of the process is $I(1)$, but where the components cannot be rewritten as Markov processes, and therefore they can not be classified in terms of null recurrence.

To summarize, according to Corollary 1, null recurrence seems to work as a generalization of (one) unit root for a scalar AR(p) process. The multivariate processes are more difficult because in general we cannot rewrite each component as a Markov process. However, parts (i) and (ii) of Theorem 4 ensure that an n-dimensional multivariate cointegrated process with $n-1$ or $n-2$ stationary relationships is null recurrent. The opposite is not correct; i.e., a process satisfying part (i) or (ii) of Theorem 4 does not necessarily cointegrate. This is so because the theorem does not exclude the processes where some of the components are stationary.

4. PROOF OF THEOREM 4

It is well known that the so-called drift conditions in Meyn and Tweedie (1993, Chs. 8 and 11) are powerful in the stationary situation, and they have been applied in the null recurrent situation also; cf., e.g., Cline and Pu (1998, 1999). However, they are often difficult to use in the null recurrent situation, and we were not able to apply them to prove Theorem 4. Besides, and crucial for us, they do not say anything about β-null recurrence; i.e., there are as yet no drift criteria for this property. Our approach is quite different and uses Theorem 1. We hope that this approach will turn out to be useful also in verifying null recurrence for nonlinear processes.

In Section 4.1 we prove Theorem 4 when the noise process is normally distributed. The general case is treated in Section 4.2.

We need four lemmas that will be applied both in the Gaussian and in the non-Gaussian cases. Lemma 3 establishes a technical detail, whereas Lemma 4 is crucial for the proof of Theorem 4. They are both easy to prove. Lemma 5, in general referred to as the Wielandt–Hoffman theorem, is a perturbation result that we use in Lemma 6 to analyze the covariance structure of the process.

If $\{X_t; t \geq 0\}$ is a VAR(1) process, i.e.,

$$X_t = AX_{t-1} + b + \varepsilon_t, \tag{4.1}$$

the m-step process $\{X_t^m = X_{tm}; t \geq 0\}$ satisfies the difference equation

$$X_t^m = \tilde{A}X_{t-1}^m + \tilde{b} + \tilde{\varepsilon}_t, \tag{4.2}$$

where the noise process $\tilde{\varepsilon}_t = \sum_{k=0}^{m} A^k \varepsilon_{t-k}$ for $t \geq m$, $\tilde{b} = \sum_{k=0}^{m} A^k b$, and $\tilde{A} = A^m$.

LEMMA 3. *Let* $\{X_t\}$ *be a* VAR(1) *Markov process of dimension* n. *Assume in addition that* $\{X_t\}$ *is* φ*-irreducible where* φ *is absolutely continuous with respect*

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

to the Lebesgue measure l on R^n. Then there is an m_1 such that all m-step chains $\{X_t^m = X_{tm}; t \geq 0\}$ with $m \geq m_1$ have a noise process $\{\tilde{\varepsilon}_t\}$ with positive definite covariance matrices.

Proof. Let $\{X_t\}$ and $\{X_t^m\}$ be defined by (4.1) and (4.2), respectively. Define $\Sigma_m = \mathrm{E}[\tilde{\varepsilon}_t \tilde{\varepsilon}_t'] = \sum_{k=0}^{m-1} A^k \Sigma (A^k)'$, where $\Sigma = \mathrm{E}[\varepsilon_t \varepsilon_t']$ for $t \geq 1$. Let N_m be the null space of Σ_m, i.e., $N_m = \{a : a \Sigma_m a' = 0\}$. Then $N_{m+1} \subset N_m$ because Σ_m is a sum of nonnegative definite matrices. Let $N_\infty = \cap_{m \geq 1} N_m$ and let $d_m = \dim(N_m)$. Now, d_m is a nonincreasing integer sequence, and $d_m \geq 0$ for $m \geq 1$. Define $d_\infty = \lim_{m \to \infty} d_m$; then either (a) or (b), which follow, applies:

(a) $d_\infty = 0$, which implies that there is an m_1 such that $d_{m_1} = 0$, and N_m is empty for all $m \geq m_1$. Therefore Σ_m has full range and is positive definite for $m \geq m_1$.
(b) $d_\infty \geq 1$. Then N_∞ is nontrivial.

We will demonstrate that (b) leads to a contradiction. Let $X_0 = 0$; then $X_t = \sum_{k=0}^{t-1} A^k b + \sum_{k=0}^{t-1} A^k \varepsilon_{t-k}$. Notice that by using t in place of m, we have that $\Sigma_t = \mathrm{Cov}(X_t)$. Now, let $a \in N_\infty$ be the first row of an orthonormal matrix R and define $Y_t = R X_t$ for $t \geq 0$. Then the first component of $\{Y_t\}$ has zero variance for $t \geq 0$. This means that Y_t, and hence X_t, cannot reach all sets with positive φ-measure. More precisely, let y_1 be the first component of an n-dimensional vector y and define the sets $\tilde{F}_t = \{y \in R^n : y_1 \neq a \sum_{k=0}^{t-1} A^k b\}$, for $t \geq 1$, and $\tilde{F} = \cup_{t \geq 0} \tilde{F}_t$. Let F be the inverse image of the set $\tilde{F}$ under the transformation R. Because $\varphi \ll l$ and $l(F^c) = 0$, we have that $\varphi(F) > 0$. But

$$\mathrm{P}_0(X_t \in F) = \mathrm{P}_0(Y_t \in \tilde{F}) = 0,$$

and this is in conflict with $\{X_t\}$ being φ-irreducible. Thus, $d_\infty = 0$. ∎

By a compact ball $C \in R^n$ we will mean a set of the form

$$C = \{x \in R^n : ||x|| \leq d_C/2\},$$

where the norm is the euclidean norm and d_C is the diameter of the ball. So in this paper, a ball is centered at zero.

LEMMA 4. *Let $\{X_t; t \geq 1\}$ be an n-dimensional stochastic process, $n \geq 1$. Assume that $X_t \overset{\mathcal{L}}{=} \mathcal{N}(\mu_t, \Sigma_t)$ where Σ_t is nonsingular for $t \geq 1$. Let $\lambda_{1,t}, \ldots, \lambda_{n,t}$ be the eigenvalues of Σ_t for $t \geq 1$. Assume that*

$$\lim_{t \to \infty} \lambda_{i,t} = \infty, \qquad i = 1, \ldots, d,$$

where $\lim_{t \to \infty} \lambda_{i,t} = \lambda_i > 0$ and finite for $i = d+1, \ldots, n$. Let $C \in R^n$ be an n-dimensional compact ball.

(i) Then there is a positive constant c such that

$$\mathrm{P}(X_t \in C) \leq c \left(\prod_{i=1}^{d} \lambda_{i,t} \right)^{-1/2} \quad \text{for } t \geq 1.$$

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

(ii) If the sequence $\mu_t' \Sigma_t^{-1} \mu_t$ *converges as* $t \to \infty$*, then*

$$\mathrm{P}(X_t \in C) \sim c_1 \left(\prod_{i=1}^{d} \lambda_{i,t}\right)^{-1/2},$$

where c_1 *is some positive constant.*

(iii) Let $\mu_t = (\mu_{1,t}, \ldots, \mu_{n,t})'$ *and assume that for every* $t \geq 1$*, there is an* $i = 1, \ldots n$ *such that*

$$|\mu_{i,t}\, \lambda_{i,t}^{-1/2}| \geq c_2 t^{1/2}, \tag{4.3}$$

where c_2 *is a positive constant independent of* t*. Then there is an* $r > 1$ *such that*

$$\mathrm{P}(X_t \in C) \leq r^{-t} \quad \textit{for } t \geq 1.$$

Proof. Because C is a ball, it is invariant to any orthonormal transformation; i.e., if R is an $n \times n$ orthonormal matrix and $x \in C$ then $Rx \in C$. Let R_t be the orthonormal matrix that diagonalizes Σ_t and let $\Lambda_t = R_t \Sigma_t R_t'$. Let $Y_t = R_t X_t$. Because C is invariant to orthonormal transformations, we have that

$$\mathrm{P}(X_t \in C) = \mathrm{P}(Y_t \in C), \tag{4.4}$$

and $\mathrm{P}(Y_t \in C)$ is equal to

$$\left(\prod_{i \leq d} \lambda_{i,t}\right)^{-1/2} \int_C (2\pi)^{-n/2} \left(\prod_{i > d} \lambda_{i,t}\right)^{-1/2} \exp\{-(y - \tilde{\mu}_t)' \Lambda_t^{-1} (y - \tilde{\mu}_t)/2\} dy, \tag{4.5}$$

where $\tilde{\mu}_t = R_t \mu_t$. The integral in (4.5) is always less than

$$c_t = \int_C (2\pi)^{-n/2} \left(\prod_{i > d} \lambda_{i,t}\right)^{-1/2} dy,$$

and by assumption c_t converges as $t \to \infty$. Therefore there is a constant c such that

$$\mathrm{P}(X_t \in C) \leq c(\lambda_{1,t} \ldots \lambda_{d,t})^{-1/2} \quad \text{for } t \geq 1,$$

and so part (i) of the lemma is proved.

To prove part (ii), note that by assumptions, Λ_t^{-1} and $\tilde{\mu}_t' \Lambda_t^{-1} \tilde{\mu}_t = \mu_t' \Sigma_t^{-1} \mu_t$ converge as $t \to \infty$. And therefore obviously $\Lambda_t^{-1} \tilde{\mu}_t$ converges as $t \to \infty$. Thus, the integral in (4.5) converges to a constant c_1. That is,

$$\mathrm{P}(X_t \in C) \sim c_1 (\lambda_{1,t} \ldots \lambda_{d,t})^{-1/2},$$

and so part (ii) of the lemma is proved.

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

Proof of part (iii). To keep the notation simple let C be a ball with radius one. Let $Y_{i,t}$ be component i of the vector $Y_t = R_t X_t$. Then

$$P(Y_t \in C) \leq P(-1 \leq Y_{i,t} \leq 1) = P\left(\frac{-1-\mu_{i,t}}{\sqrt{\lambda_{i,t}}} \leq U \leq \frac{1-\mu_{i,t}}{\sqrt{\lambda_{i,t}}}\right), \tag{4.6}$$

where U is $\mathcal{N}(0,1)$ distributed. By assumption $\lambda_{i,t}^{-1/2}$ is bounded, and so by (4.6) and assumption (4.3) there is an $r > 1$ such that

$$\mathrm{P}(Y_t \in C) < r^{-t}, \tag{4.7}$$

and part (iii) follows by (4.4) and (4.7). ■

The following lemma, which in general is referred to as the Wielandt–Hoffman theorem, can be found in Stewart and Sun (1990) and is crucial for discussing the covariance structure of the VAR(1) process.

First we define the Frobenius norm $||\cdot||_F$ of an $n \times m$ matrix A,

$$||A||_F = \sqrt{\textstyle\sum_{i=1}^{n}\sum_{j=1}^{m}|a_{i,j}|},$$

where $a_{i,j}$, $i = 1,\ldots,n$, $j = 1,\ldots,m$, are the elements of A. Note that a matrix $S \in \mathcal{C}^{n\times n}$ is called normal if $SS^* = S^*S$ where S^* means complex conjungate transposed, and most important for us it also means that any Hermitian matrix is normal.

LEMMA 5 (Wielandt–Hoffman). *Let S and $\Sigma = S + E$ be normal $n \times n$-matrices. Then*

$$\sum_{i=1}^{n}|\lambda_i - \rho_i| \leq ||E||_F,$$

where $\lambda_1 \geq \cdots \geq \lambda_n$ and $\rho_1 \geq \cdots \geq \rho_n$ are the eigenvalues of Σ and S, respectively.

Recall from Section 2 that for the expectation operator, $\mathrm{E}_0 = \mathrm{E}_{\delta_0}$ where δ_0 is the point measure concentrated at zero.

LEMMA 6. *Let $\{X_t\}$ be an φ-irreducible* VAR(1) *process of dimension $n \geq 1$ as defined in (3.1), where φ is absolutely continuous with respect to the Lebesgue measure. Assume that the eigenvalues of the matrix A are on or inside the unit circle and that the matrix B_1 defined in (3.2) is diagonal of dimension $d \times d$, $d \geq 1$.*

Let $\Sigma_t^X = \mathrm{Cov}(X_t|X_0 = 0)$ and let $\lambda_{1,t} \geq \cdots \geq \lambda_{n,t}$ be the eigenvalues of Σ_t^X for $t \geq 1$. Then

$$\lambda_{i,t} \sim t c_i, \qquad i = 1,\ldots,d, \tag{4.8}$$

as $t \to \infty$, where c_i are positive constants, $i = 1,\ldots,d$. And when $d < n$

$$\lim_{t\to\infty}\lambda_{i,t} = \lambda_i > 0 \qquad \text{for } i = d+1,\ldots,n. \tag{4.9}$$

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

Assume in addition that the constant vector $b \in G_A$. Let $\mu_t^X = \mathrm{E}_0[X_t]$. Then

$$\lim_{t\to\infty} (\mu_t^X)'(\Sigma_t^X)^{-1}\mu_t^X = c \geq 0. \tag{4.10}$$

Assume that $b \notin G_A$. Let $\mu_t^X = [\mu_{1,t}^X, \ldots, \mu_{n,t}^X]'$. Then there is a constant $c_2 > 0$ such that for every $t \geq 1$, there is an $i \in \{1, \ldots n\}$, so that

$$|\mu_{i,t}^X| \geq c_2 t \quad \textit{and} \quad (\mu_{i,t}^X)^2 \lambda_{i,t}^{-1} \geq c_2 t. \tag{4.11}$$

Proof. Let A be an $n \times n$ dimensional matrix with d eigenvalues on the unit circle. Write

$$X_t = A^t X_0 + \sum_{k=0}^{t-1} A^k b + \sum_{k=0}^{t-1} A^k \varepsilon_{t-k}.$$

We divide the proof into two steps.

Step 1. The covariance when B_1 in (3.2) is diagonal. In the first step we analyze the covariance structure of X_t, as t goes to infinity, when B_1 is diagonal (see (3.2) for the definition of B_1 and B_2). Let

$$\Sigma_t^X = \mathrm{Cov}(X_t | X_0 = 0).$$

First, recall that every m-step chain is β-null recurrent if and only if $\{X_t\}$ is; see Lemma 2. So by Lemma 3 we assume without loss of generality that the covariance matrix $\Sigma_\varepsilon = \mathrm{E}[\varepsilon_t \varepsilon_t{}']$ is positive definite.

Recall that $A^k = Q B^k Q^{-1}$ where B is the Jordan form and consists of the blocks B_1 and B_2 introduced in (3.2) and Q is a nonsingular matrix.
Define

$$Y_t = Q^{-1} X_t, \qquad \eta_t = Q^{-1}\varepsilon_t, \qquad b_Y = Q^{-1} b,$$

so that

$$Y_t = B Y_{t-1} + b_Y + \eta_t.$$

Let

$$Y_t = \begin{bmatrix} Y_t^1 \\ Y_t^2 \end{bmatrix}, \qquad \eta_t = \begin{bmatrix} \eta_t^1 \\ \eta_t^2 \end{bmatrix}, \qquad b_Y = \begin{bmatrix} b_Y^1 \\ b_Y^2 \end{bmatrix}, \qquad B = \begin{bmatrix} B_1 & 0 \\ 0 & B_2 \end{bmatrix}, \tag{4.12}$$

where Y_t^1, η_t^1, and b_Y^1 have dimension d and Y_t^2, η_t^2, and b_Y^2 have dimension $(n-d)$. Then

$$Y_t^i = B_i Y_{t-1}^i + b_Y^i + \eta_t^i \quad \text{for } i = 1, 2.$$

Let $\Sigma_t = \mathrm{Cov}(Y_t | Y_0 = 0)$ and $\Sigma_{ij,t} = \mathrm{Cov}(Y_t^i, Y_t^j | Y_0 = 0)$ for $i, j = 1, 2$.

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

We have that $\Sigma_t = \sum_{k=0}^{t-1} H^{(k)}$, where

$$H^{(k)} \stackrel{\text{def}}{=} \text{Cov}(B^k \eta_{t-k}) = B^k V B^{k^*}, \qquad 0 \le k < t, \tag{4.13}$$

where $V = Q^{-1} \Sigma_\varepsilon (Q^*)^{-1}$ and Q^* and B^{k*} are conjugates transposed of Q and B^k, respectively. Write the Hermitian matrix $H^{(k)}$ in the following way:

$$H^{(k)} = \begin{bmatrix} H_{11}^{(k)} & H_{12}^{(k)} \\ H_{21}^{(k)} & H_{22}^{(k)} \end{bmatrix} = \begin{bmatrix} B_1^k V_{11} B_1^{k^*} & B_1^k V_{12} B_2^{k^*} \\ B_2^k V_{21} B_1^{k^*} & B_2^k V_{22} B_2^{k^*} \end{bmatrix}.$$

Because the diagonal elements of B_2 are less than one, we find that for $(i, j) \neq (1, 1)$,

$$\Sigma_{ij,t} = \Sigma_{ij} + \Delta_{ij}(t), \qquad \Delta_{ij}(t) = o(r^t) \quad \text{for some } r \in (0, 1), \tag{4.14}$$

where Σ_{ij} are matrices of constant elements independent of t.

Let $h_{ij}^{(k)}$, $i, j = 1, \ldots, n$, denote the elements of the matrix $H^{(k)}$ for $k \ge 1$. When B_1 is diagonal, all the elements of B_1^k are on the unit circle, and they can be written as $e^{ik\theta_i}$ where $-\pi < \theta_i \le \pi$. The elements $\sigma_{ij}^{\gamma}(t)$ of $\Sigma_{11,t}$ are power series and are given by

$$\sigma_{ij}^{\gamma}(t) = \sum_{k=0}^{t-1} h_{ij}^{(k)} = v_{ij} \sum_{k=0}^{t-1} e^{ik(\theta_i - \theta_j)} = \begin{cases} t v_{ij}, & \text{if } \theta_i = \theta_j \\ O(1), & \text{otherwise,} \end{cases} \tag{4.15}$$

as $t \to \infty$, and where v_{ij}, $i, j = 1, \ldots, d$, are the elements of the matrix V_{11}. Thus, we may write

$$\Sigma_{11,t} = t\mathcal{D}_{11} + \Delta_{11}(t), \quad \text{where } \mathcal{D}_{11} = \lim_{t \to \infty} t^{-1} \sum_{k=0}^{t-1} H_{11}^{(k)}, \qquad \Delta_{11}(t) = O(1). \tag{4.16}$$

So element (i, j) of the matrix $\mathcal{D}_{11}$ is equal to element $v_{i,j}$ of the matrix V_{11} when $\theta_i = \theta_j$, and otherwise it is zero, for $i, j = 1, \ldots, d$; see (4.15). That is, $\text{diag}(\mathcal{D}_{11}) = \text{diag}(V_{11})$, and the off-diagonal elements are zero if B_1 has only distinct elements. But if B_1 has two or more equal elements some of the off-diagonal elements of $\mathcal{D}_{11}$ are nonzero and equal to the corresponding elements of V_{11}.

The matrix Σ_ε is assumed to be positive definite, and Q is nonsingular; therefore V is positive definite. So, by (4.15), $\mathcal{D}_{11}$ is positive definite also. To see this, define

$$M_i^{(j)} = \begin{bmatrix} v_{i,i} & \cdots & v_{i,(i+j)} \\ \vdots & & \vdots \\ v_{(i+j),i} & \cdots & v_{(i+j),(i+j)} \end{bmatrix} \quad \text{for } i = 1, \ldots, n, \qquad j = 0, \ldots, n-i,$$

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

and note that $M_i^{(j)}$, $i = 1, \ldots, n$, and $j = 0, \ldots, n-i$, are positive definite if V is. So, by (4.15), $\mathcal{D}_{11}$ is positive definite.

When B_1 is diagonal, this gives the following representation of the covariance matrix $\Sigma_t^X = \mathrm{Cov}(X_t | X_0 = 0)$ for $t \geq 1$:

$$\Sigma_t^X = tQ \begin{bmatrix} \mathcal{D}_{11} & 0 \\ 0 & 0 \end{bmatrix} Q^* + Q \begin{bmatrix} \Delta_{11}(t) & \Sigma_{12} + \Delta_{12}(t) \\ \Sigma_{21} + \Delta_{21}(t) & \Sigma_{22} + \Delta_{22}(t) \end{bmatrix} Q^*. \tag{4.17}$$

And we can write

$$\Sigma_t^X = t\mathcal{D} + O(1), \tag{4.18}$$

where the matrix $\mathcal{D}$ is the matrix in the first part of (4.17). Applying Lemma 5 to (4.18) we find that the differences between the d largest eigenvalues $\lambda_{1,t}, \ldots, \lambda_{d,t}$ of Σ_t^X and the d eigenvalues of $t\mathcal{D}$ are bounded; i.e.,

$$\lambda_{i,t} \sim tc_i, \qquad i = 1, \ldots, d. \tag{4.19}$$

Now we will show that the $n - d$ smaller eigenvalues converge to positive numbers as $t \to \infty$. Let $F_t = \Sigma_t^{-1}$. Then the block matrices of F_t are

$$F_{11}(t) = (\Sigma_{11,t} - \Sigma_{12,t}\Sigma_{22,t}^{-1}\Sigma_{21,t})^{-1},$$

$$F_{12}(t) = -(\Sigma_{11,t} - \Sigma_{12,t}\Sigma_{22,t}^{-1}\Sigma_{21,t})^{-1}\Sigma_{12,t}\Sigma_{22,t}^{-1} \quad \text{and} \quad F_{21}(t) = (F_{12}(t))^*,$$

$$F_{22}(t) = (\Sigma_{22,t} - \Sigma_{21,t}\Sigma_{11,t}^{-1}\Sigma_{12,t})^{-1}.$$

Let $F_t^X = (Q^*)^{-1} F_t Q^{-1}$, i.e., F_t^X is the inverse matrix to Σ_t^X. Moreover, let

$$F_{1,t}^X = (Q^*)^{-1} \begin{bmatrix} 0 & 0 \\ 0 & F_{22}(t) \end{bmatrix} Q^{-1} \quad \text{and} \quad F_{2,t}^X = (Q^*)^{-1} \begin{bmatrix} F_{11}(t) & F_{12}(t) \\ F_{21}(t) & 0 \end{bmatrix} Q^{-1}.$$

So $F_t^X = F_{1,t}^X + F_{2,t}^X$. By (4.16) we find that $\lim_{t\to\infty} F_{2,t}^X = 0$, the zero matrix. Thus by Lemma 5 the eigenvalues of F_t^X differ only slightly from the eigenvalues of $F_{1,t}^X$. Of course this means that d eigenvalues of F_t^X converge to zero. Moreover it means that whether the $n - d$ last eigenvalues of F_t^X converge or not is due to $F_{1,t}^X$, which in turn is due to whether the eigenvalues of (see (4.14))

$$(F_{22}(t))^{-1} = \Sigma_{22,t} - \Sigma_{21,t}\Sigma_{11,t}^{-1}\Sigma_{12,t} = \Sigma_{22} + (\Delta_{22}(t) - \Sigma_{21,t}\Sigma_{11,t}^{-1}\Sigma_{12,t})$$

converge or not. By (4.14) and (4.16) we see that $\lim_{t\to\infty}(\Delta_{22}(t) - \Sigma_{21,t}\Sigma_{11,t}^{-1}\Sigma_{12,t}) = 0$, the zero matrix. Therefore by Lemma 5 the eigenvalues of $(F_{22}(t))^{-1}$ converge to the eigenvalues of Σ_{22} which are positive. So the $n - d$ smallest eigenvalues, $\lambda_{d+1,t}, \ldots, \lambda_{n,t}$, of Σ_t^X converge; i.e.,

$$\lim_{t\to\infty} \lambda_{i,t} = \lambda_i \quad \text{for } i = d+1, \ldots, n. \tag{4.20}$$

So (4.8) and (4.9) of Lemma 6 are satisfied by (4.19) and (4.20), respectively.

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

Note that

$$\lim_{t\to\infty} \Sigma_t^{-1} = \lim_{t\to\infty} F_t = \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix}, \tag{4.21}$$

which is to be used in the next step. This holds because, by letting $\tilde{\Delta}(t) = \Delta_{22}(t) - \Sigma_{12,t}^{*}\Sigma_{11,t}^{-1}\Sigma_{12,t}$, we have that $\Sigma_{22}^{-1}(\Sigma_{22} + \tilde{\Delta}_{22}(t)) = I + \Sigma_{22}^{-1}\tilde{\Delta}_{22}(t)$, and the elements of the last term go to zero.

Step 2. The quadratic form and the mean when B_1 is diagonal. We will study the quadratic form $(\mu_t^X)'(\Sigma_t^X)^{-1}\mu_t^X$. Start the chain at $X_0 = 0$ and, as stated in Lemma 6, recall that

$$\mu_t^X = \mathrm{E}_0[X_t] = \sum_{k=0}^{t-1} A^k b, \qquad t \geq 1.$$

Let the matrices Q, B, Σ_t^X, Σ_t, and $\Sigma_{ij}(t)$, $i, j = 1, 2$, be as in step 1 of the proof. Let $\gamma_1, \ldots, \gamma_n$ be the eigenvalues of the matrix A such that $|\gamma_1| \geq \cdots \geq |\gamma_n|$. That is, γ_j is diagonal element j of the matrix B.

The diagonal elements γ_j, $j+1, \ldots, d$, of B_1 are on the unit circle; i.e., they can be written as $\gamma_j = e^{ik\theta_j}$ where $-\pi < \theta_j \leq \pi$. The power series give that

$$\gamma_{j,t} \stackrel{\text{def}}{=} \sum_{k=0}^{t-1} \gamma_j^k = \sum_{k=0}^{t-1} e^{ik\theta_j} = \begin{cases} t, & \text{if } \theta_j = 0 \\ O(1), & \text{otherwise.} \end{cases} \tag{4.22}$$

Let $\mu_t = Q^{-1}\mu_t^X$ and recall by (4.12) that $b_Y = Q^{-1}b$. Note that

$$b_Y = \alpha = [\alpha_1, \ldots, \alpha_n]',$$

where α is the complex vector defined in (3.4). Then

$$\mu_t = \sum_{k=0}^{t-1} B^k b_Y, \qquad \mu_t^1 = \sum_{k=0}^{t-1} B_1^k b_Y^1, \quad \text{and} \quad \mu_t^2 = \sum_{k=0}^{t-1} B_2^k b_Y^2 \quad \text{where } \mu_t = \begin{bmatrix} \mu_t^1 \\ \mu_t^2 \end{bmatrix}.$$

Note that μ_t^2 converges as $t \to \infty$.

Recall by (3.5) that $b \in G_A$ means that if $\gamma_i = 1$, then $\alpha_i = 0$ for every $i = 1, \ldots, k$ where $k \leq d$. So by (4.22),

$$\mu_t^1 = O(1) \quad \text{when } b \in G_A, \tag{4.23}$$

where $O(1)$ means a vector with bounded elements. Moreover, by (4.22), there is at least one element $\mu_{i,t}$, $i = 1, \ldots, d$, of μ_t^1 such that

$$\mu_{i,t} = \alpha_i t, \qquad \alpha_i \neq 0, \quad \text{when } b \notin G_A. \tag{4.24}$$

Therefore (4.11) of the lemma follows by (4.8), (4.9), and (4.24).

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

Recall by (4.21) that

$$\lim_{t\to\infty} \Sigma_t^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix}. \tag{4.25}$$

We have that

$$(\mu_t^X)'(\Sigma_t^X)^{-1}\mu_t^X = \mu_t^* \Sigma_t^{-1} \mu_t. \tag{4.26}$$

If $b \in G_A$ it follows by (4.23) and (4.25) that

$$\lim_{t\to\infty} \mu_t^\star \Sigma_t^{-1} \mu_t = (\mu^2)^* \Sigma_{22}^{-1} \mu^2 = c, \tag{4.27}$$

where $c \geq 0$ is a constant. Thus (4.10) of the lemma follows by (4.26) and (4.27). ■

4.1. Proof of Theorem 4 in the Gaussian Case

We will prove the five points (i)–(v) of Theorem 4, under the assumption of a normally distributed noise process $\{\varepsilon_t\}$. Recall that by Lemma 2 we can work with the m-step chain without loss of generality, and in what follows we do so without further comment.

(i) In this case $d = 1$, i.e., $B_1 = 1$ in (3.2). Lemma 6 ensures that the conditions in part (ii) of Lemma 4 hold. Thus there exist a small set C and a constant c such that

$$\mathrm{P}_0(X_t \in C) \sim ct^{-1/2}.$$

So $\{X_t\}$ is β-null recurrent with $\beta = \frac{1}{2}$ by Lemma 1 and (2.13).

(ii) Here $d = 2$. Again, Lemma 6 and part (ii) of Lemma 4 imply that there are a small set C and a constant c such that

$$\mathrm{P}_0(X_t \in C) \sim ct^{-1}.$$

Thus $\sum_{t=0}^{\infty} \mathrm{P}_0(X_t \in C) \equiv \infty$, which means that $\{X_t\}$ is recurrent (see Theorem 1). It is null recurrent by Lemma 1. Because $\sum_{t=0}^{T} t^{-1} \sim \ln(T)$, there is no $\beta \in (0, 1)$ such that Definition 3 is satisfied.

(iii) First, let J be a $p \times p$, $2 \leq p \leq d$, Jordan matrix with the eigenvalue $e^{i\theta}$ where $-\pi < \theta \leq \pi$. And let $J_{l,j}^{(k)}$ be element (l, j) of J^k for $l, j = 1, \ldots, p$ and $k \geq 1$. Then straightforward calculations show that for $k \geq 1$,

$$J_{l,j}^{(k)} = \begin{cases} 0, & \text{for } l < j, \\ e^{ik\theta}, & \text{for } l = j, \\ (k - l + j + 1)e^{i(k-l+j)^+\theta}, & \text{for } l > j, \end{cases} \tag{4.28}$$

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

where $(x)^+ = \max(0, x)$. That is, the absolute value of any element in the lower triangle of the matrix J^k increases like k. So, e.g., if B_1 is a two-dimensional elementary Jordan matrix with eigenvalue one, then $b_{2,1}^{(k)} = k$ where $b_{2,1}^{(k)}$ is element $(2, 1)$ of B_1^k. Thus some diagonal elements of $\mathrm{Cov}(A^k \varepsilon_{t-k})$, increase like k^2, and

$$t^{-3} \Sigma_t^X = t^{-3} \mathrm{Cov}(X_t | X_0 = 0) \tag{4.29}$$

converges to a matrix with some positive elements. Thus Σ_t^X has at least one eigenvalue $\lambda_{1,t}$ which increases like t^3. Therefore, by part (i) of Lemma 4, $\mathrm{P}_0(X_t \in C)$ goes to zero at least as fast as $t^{-3/2}$. Hence $\sum_{t=0}^{\infty} \mathrm{P}_0(X_t \in C)$ is finite, and Theorem 1 ensures that $\{X_t\}$ is transient.

Note that these arguments are independent of the mean μ_t^X, i.e., the conclusion holds for every constant vector b in (3.1).

(iv) Suppose that B_1 has dimension $d \geq 3$ but does not contain an elementary Jordan matrix, i.e., B_1 is diagonal. Then Lemma 6 and part (i) of Lemma 4 imply that for some constant c,

$$\mathrm{P}_0(X_t \in C) \leq c t^{-d/2}, \qquad d \geq 3.$$

Therefore $\sum_{t=0}^{\infty} \mathrm{P}_0(X_t \in C)$ converges, and according to Theorem 1, $\{X_t\}$ is transient.

If B_1 has dimension $d \geq 3$ and contains an elementary Jordan matrix, the process is transient. It clearly follows by the same argument as in point (iii). Note that these arguments hold for every constant vector b in (3.1).

(v) When B_1 is nondiagonal the result is proved in part (iii). So we only have to consider the case where B_1 is diagonal and $b \notin G_A$. In that case Lemma 6 provides the assumptions of part (iii) of Lemma 4, which ensures that $\sum_{t=0}^{\infty} \mathrm{P}_0(X_t \in C)$ converges. Thus, $\{X_t\}$ is transient by Theorem 1. ■

4.2. Proof of Theorem 4 in the Non-Gaussian Case

Lemma 4, which we applied to prove Theorem 4 in the normal case, will be used in the nonnormal case also but with normality replaced by asymptotic normality. To apply Lemma 4 we need a local central limit theorem for densities. However, this is not straightforward to derive because the densities of X_t, $t \geq 1$, where $\{X_t\}$ is a VAR(1) process, do not in general possess a local limit theorem. That is, in general the densities, q_t, of $t^{-1/2} X_t$ do not satisfy

$$\sup_x |q_t(x) - \phi_\Gamma(x)| \to 0 \quad \text{as } t \to \infty, \tag{4.30}$$

where ϕ_Γ is a normal distribution density with some nonsingular covariance matrix Γ. A stationary VAR(1) process whose noise generating process is non-normally distributed is an example of this. But if $\{X_t\}$ is a random walk,

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

$t^{-1/2}X_t$ is asymptotically normal, and (4.30) is true under appropriate conditions. Lemma 10, which appears later in this section, deals with this problem in the case of a VAR(1) process with at least one root on the unit circle and the other roots outside the unit circle.

The approach in Lemma 10 consists of two parts. Suppose that the process $\{X_t\}$ can be written as a process $\{(Z_t', V_t')'\}$ where speaking informally $\{V_t\}$ is a stationary process, or more precisely it is an ergodic process, i.e., (2.2) holds. Moreover, $\{Z_t\}$ is a "nonstationary" process such that speaking informally, the components are unit root processes, but they are not cointegrated. This will be made precise subsequently. The idea is that the "nonstationary" and the "stationary" components are asymptotically independent. More exactly, we show that asymptotically the "stationary" component is determined by its recent history, whereas (compared to the whole history) the recent history is asymptotically negligible for the "nonstationary" component. So essentially, the "stationary" and the "nonstationary" components are determined by different segments of the history.

To separate the process $\{X_t\}$ into a d-dimensional random walk and an $(n-d)$ dimensional stationary component, Schur's triangularization theorem (see Horn and Johnson, 1990, p. 79) is helpful, and we state it here for the reader's convenience.

LEMMA 7. *If A is an $n \times n$ matrix with real elements, there is an $n \times n$ real orthogonal matrix $\tilde{Q}$ such that*

$$\tilde{Q}' A \tilde{Q} = \begin{bmatrix} A_1 & & \star \\ & \ddots & \\ 0 & & A_k \end{bmatrix}, \tag{4.31}$$

where each A_i is a real 1×1 matrix, consisting of a single eigenvalue, or a real 2×2 matrix with a nonreal pair of complex conjugate eigenvalues. The 0 means that all the elements are zero, and the $\star$ denotes any entry of real elements. The diagonal blocks may be arranged in any prescribed order.

Assume that the $n \times n$ matrix A has d eigenvalues on the unit circle and $(n-d)$ eigenvalues inside the unit circle. It is convenient for us to write (4.31) in the form

$$\tilde{Q}' A \tilde{Q} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \tag{4.32}$$

where the $d \times d$ matrix A_{11} has all the eigenvalues on the unit circle and the $(n-d) \times (n-d)$ matrix A_{22} has all the eigenvalues inside the unit circle.

By Theorem 2, we may assume without loss of generality that the matrix A has the form of the right-hand side of (4.32). It means that the d first components of $\{X_t\}$ are the nonstationary part, whereas the $n-d$ last components are the stationary part. The stationary part and the nonstationary part are dependent in general.

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

Let $\mu_t^X = \mathrm{E}_0[X_t]$ and let μ_t^Z and μ_t^V consist of the d first components of μ_t^X and the $n-d$ last components of μ_t^X, respectively. Let

$$\tilde{\mu}_t^Z = \begin{bmatrix} \mu_t^Z \\ 0 \end{bmatrix} \quad \text{and} \quad \tilde{\mu}_t^V = \begin{bmatrix} 0 \\ \mu_t^V \end{bmatrix}, \qquad t \geq 1,$$

where the zeros mean that the last $n-d$ components and the first d components are zero, respectively, for the n-dimensional vectors $\tilde{\mu}_t^Z$ and $\tilde{\mu}_t^V$. Write

$$X_t = X_t^\star + X_{\star t}, \qquad t \geq 1, \tag{4.33}$$

where

$$X_t^\star \stackrel{\text{def}}{=} \sum_{k=m_t}^{t-1} A^k \varepsilon_{t-k} + \tilde{\mu}_t^Z, \qquad X_{\star t} \stackrel{\text{def}}{=} \sum_{k=0}^{m_t-1} A^k \varepsilon_{t-k} + \tilde{\mu}_t^V, \tag{4.34}$$

and the integer function $m_t = [c \ln(t)]$, where c is some positive constant. Moreover, A has the form of the right-hand side of (4.32). So $\{X_t^\star, t \geq 1\}$ represents the early history of the process, whereas $\{X_{\star t}, t \geq 1\}$ is the recent history. Let

$$X_t = \begin{bmatrix} Z_t \\ V_t \end{bmatrix}, \qquad X_t^\star = \begin{bmatrix} Z_t^\star \\ \Delta_t \end{bmatrix}, \quad \text{and} \quad X_{\star t} = \begin{bmatrix} \Theta_t \\ V_t^\star \end{bmatrix}, \quad t \geq 1. \tag{4.35}$$

Here Z_t, $Z_t^\star$, and Θ_t consist of the d first components of X_t, $X_t^\star$, and $X_{\star t}$, respectively, and correspond to the nonstationary part of the process $\{X_t\}$. And V_t, Δ_t, and $V_t^\star$ are the last $(n-d)$ components of X_t, $X_t^\star$, and $X_{\star t}$, respectively, and correspond to the stationary part. Note that $\mathrm{E}_0[Z_t^\star] = \mathrm{E}_0[Z_t] = \mu_t^Z$, $\mathrm{E}_0[V_t^\star] = \mathrm{E}_0[V_t] = \mu_t^V$, $\mathrm{E}_0[\Delta_t] = 0$, and $\mathrm{E}_0[\Theta_t] = 0$.

Note that the decomposition defined by (4.33)–(4.35) can be carried out in the same manner for the m-step chain.

When $b \in G_A$, our first claim, which will be proved by a local central limit theorem and Lemma 4, is that

$$\mathrm{P}_0(Z_t \in D) \sim c_D t^{-d/2}, \qquad c_D \text{ is a positive constant}, \qquad D \in \mathcal{B}^{d^+}, \tag{4.36}$$

in the case that μ_t^Z converges.

Our second claim is that the variables Δ_t and Θ_t are asymptotically negligible, whereas the major parts, i.e., $Z_t^\star$ and $V_t^\star$, are asymptotically independent. Hence, as $t \to \infty$,

$$\mathrm{P}_0(Z_t \in D, V_t \in S) = \mathrm{P}_0(Z_t \in D)\mathrm{P}_0(V_t \in S) + o(t^{-d/2}), \qquad S \in \mathcal{B}^{n-d}. \tag{4.37}$$

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

By (4.36) and (4.37) we find that $\mathrm{P}_0(Z_t \in D, V_t \in S) \sim c_{D,S} t^{-d/2}$, $c_{D,S}$ is a positive constant. We restate this, in addition to (4.37), in Lemma 10. Assumptions A1–A4, which follow, regarding the processes in (4.35), are needed in Lemma 10.

A1. For $D, C \in \mathcal{B}^d$, and $S \in \mathcal{B}^{n-d}$,

$$\mathrm{P}_0(Z_t^\star \in D, \Theta_t \in C, V_t^\star \in S) = \mathrm{P}_0(Z_t^\star \in D)\mathrm{P}_0(\Theta_t \in C, V_t^\star \in S) + o(t^{-d/2}). \tag{4.38}$$

A2. If $l(N) = 0$, then

$$\mathrm{P}_0(V_t \in N) = 0, \quad N \in \mathcal{B}^{n-d}, \tag{4.39}$$

where l is the Lebesgue measure on R^{n-d}. Furthermore, there is a probability measure π such that

$$\mathrm{P}_0(V_t \in \cdot) \to \pi(\cdot) \quad \text{for } t \to \infty. \tag{4.40}$$

A3. For some sequence $\delta_t = o(1)$,

$$\mathrm{P}_0(||\Delta_t|| > \delta_t) = o(t^{-d/2}). \tag{4.41}$$

Moreover,

$$\mathrm{E}_0|\Theta_t \cdot 1| = o(t^{1/2}), \tag{4.42}$$

where 1 denotes the vector in R^d with one in each direction and $\cdot$ indicates the euclidean inner product.

A4. We can write

$$Z_t^\star = \sum_{k=1}^{T_t} \xi_k + \mu_t^Z, \tag{4.43}$$

where $\{\xi_t\}$ is a sequence of independent zero mean d-dimensional random vectors, T_t is an integer valued function of t such that $T_t \sim t$ as $t \to \infty$, and μ_t^Z is a constant vector for each $t \geq 0$.

The eigenvalues $\lambda_{1,t}, \ldots, \lambda_{d,t}$ of $\Gamma_t \stackrel{\text{def}}{=} \mathrm{Cov}(T_t^{-1/2} Z_t^\star)$ satisfy

$$\lim_{t\to\infty} \lambda_{i,t} = c_i, \qquad i = 1, \ldots, d, \tag{4.44}$$

where $c_1, \ldots, c_d$ are finite constants different from zero. Moreover,

$$\sup_t \mathrm{E}||\xi_t||^3 < \infty, \tag{4.45}$$

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

and the characteristic functions

$$g_{\xi_t}(u) = \mathrm{E}[\exp(iu \cdot \xi_t)], \qquad t \geq k,$$

satisfy, for some $k \geq 1$,

$$\sup_{t \geq k} \int |g_{\xi_t}(u)| du < \infty, \qquad \sup_{||u|| > a, t \geq k} |g_{\xi_t}(u)| < 1 \quad \text{for all } a > 0. \quad \textbf{(4.46)}$$

Assumption A3 will ensure that the variables Δ_t and Θ_t are asymptotically negligible. We need A4 to apply a local limit theorem to $Z_t^\star$.

The following lemma will be used to establish the connection between the conditions C1–C3 and Assumption A4, which will be made clear in Lemma 9 later in this section. The link is the m-step chain. For each $t \geq 1$, the d-dimensional vectors ξ_t in (4.43) consist of $d \leq n$ of the components of the noise vectors from the m-step chain. That is, each of the variables ξ_t, $t \geq 1$, in (4.43) will be constructed as a linear combination of m of the variables in the sequence $\{\varepsilon_t\}$.

LEMMA 8. *Let $\{\varepsilon_t; t \geq 1\}$ be i.i.d. vectors of dimension $n \geq 1$. Let $\tilde{\varepsilon}_t$ be the stochastic vector consisting of the nonzero elements of ε_t, $t \geq 1$. Let $\tilde{\varepsilon} = \tilde{\varepsilon}_1$ and assume that $\tilde{\varepsilon}$ satisfies C1–C3.*

Let $d \leq n$ and $A_t^{(j)}$ be matrices of dimension $d \times n$, for $j \geq 0$ and $t \geq 1$. Let $m \geq 1$ and define the d-dimensional vectors

$$\xi_t = \sum_{j=0}^{m-1} A_t^{(j)} \varepsilon_{tm-j} \quad \textit{for } t \geq 1.$$

Let $\Sigma_t^\xi = \mathrm{Cov}(\xi_t)$ be such that the eigenvalues satisfy $c \geq \rho_{1,t} \geq \cdots \geq \rho_{d,t} \geq c_1$ for all $t \geq 1$, where c and c_1 are positive constants. Then ξ_t satisfies

$$\sup_{t \geq 1} \mathrm{E}||\xi_t||^3 < \infty, \qquad \textbf{(4.47)}$$

$$\sup_{t \geq 1} \int |g_{\xi_t}(u)| du < \infty, \qquad \sup_{t \geq 1, ||u|| > a} |g_{\xi_t}(u)| < 1 \quad \textit{for all } a > 0, \qquad \textbf{(4.48)}$$

where g_{ξ_t} is the characteristic function of ξ_t.

Proof. Because $\rho_{1,t} \leq c$, (4.47) is obviously satisfied.

Construct the vector η_t such that it consists of all the nonzero elements from the vectors $\varepsilon_{(t-1)m}, \ldots, \varepsilon_{tm}$, for $t \geq 1$. Now, we can write

$$\xi_t = H_{d,t} \eta_t,$$

where $H_{d,t}$ is a $d \times p$ matrix with $p = mr$ and r is the number of nonzero elements of the vector ε_1. Note that $d \leq p$ because Σ_1^ξ is nonsingular.

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

Because the random vectors ε_t, $t \geq 1$, are independent, the characteristic function, $g_\eta(u)$, of the vector $\eta = \eta_1$ is absolutely integrable by the assumptions. Thus, if H is any nonsingular $p \times p$ matrix, the characteristic function of the vector $H\eta$ is absolutely integrable too. It is so because

$$g_{H\eta}(u) = g_\eta(Hu) \quad \text{and} \tag{4.49}$$

$$\int |g_{H\eta}(u)|du = \int |g_\eta(Hu)|du = |H^{-1}| \int |g_\eta(u)|du < \infty. \tag{4.50}$$

The $d \times p$ matrices $H_{d,t}$, $t \geq 1$, can be extended to a nonsingular $p \times p$ matrix H_t that satisfies

$$c_3 \leq |H_t| \leq c_4 \quad \text{for } t \geq 1, \tag{4.51}$$

where c_3 and c_4 are positive constants. Therefore, by using H_t in place of H in (4.49) and (4.50), it follows by (4.51) that

$$\sup_t \int g_{\xi_t}(u)du < \infty; \tag{4.52}$$

i.e., the first part of (4.48) is satisfied.

By Condition C.3 it follows that

$$\sup_{||u||>a} |g_\eta(u)| < 1 \quad \text{for all } a > 0.$$

Thus, the second part of (4.48) follows by (4.51) and (4.49) with H_t in place of H. ■

LEMMA 9. *Let $\{X_t\}$ be a φ-irreducible* VAR(1) *process defined by (3.1) satisfying C1–C3, where φ is absolutely continuous with respect to the Lebesgue measure. Assume that the eigenvalues of the matrix A are on or inside the unit circle and the matrix B_1 defined in (3.2) is diagonal of dimension $d \times d$, $d \geq 1$.*

For the m-step chain, let $Z_t^\star$, $V_t^\star$, Θ_t, and Δ_t be the decomposition of X_t^m, $t \geq 1$, defined explicitly in (4.35) for $m = 1$. Then there is a finite m such that Assumptions A1–A4 are satisfied.

Proof. First, recall by (4.2) that $\{X_t^m\}$ is a VAR(1) process with i.i.d. errors, for any $m \geq 1$. However, to keep the notation simple we will for the most part work with $m = 1$ in this proof.

Verifying Assumption A1: Clearly $Z_t^\star$ is independent of both $V_t^\star$ and Θ_t, and so the assumption stated in equation (4.38) of A1 is satisfied for any m-step chain.

Verifying Assumption A3: Recall that V_t is the $(n-d)$ last component of X_t and let $\Sigma_{22}^{\mathcal{E}}$ be the $(n-d) \times (n-d)$ covariance matrix corresponding to the $(n-d)$ last component of ε_1. We can assume without loss of generality that A has the form of the right-hand side of (4.32). Therefore $\mathrm{Cov}(V_t) = \sum_{k=0}^{t-1} A_{22}^k \Sigma_{22}^{\mathcal{E}} (A_{22}^k)'$.

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

So $\mathrm{Cov}(V_t)$ converge geometrically fast as $t \to \infty$. That is, there are finite constants K and M and an $r > 1$ such that every element of $\mathrm{Cov}(V_t)$ is less than $K \sum_{k=0}^{t-1} r^{-k} \leq KM$ for every $t \geq 0$. Therefore, and because $\mathrm{Cov}(\Delta_t) = \sum_{k=m_t}^{t-1} A_{22}^k \Sigma_{22}^{\varepsilon} (A_{22}^k)'$, we find that $\mathrm{Var}(||\Delta_t||) \leq \sqrt{d}\, KMr^{-c\ln(t)}$ for all $t \geq 0$, from which it is obvious that we can choose c such that

$$\mathrm{Var}(||\Delta_t||) = o(t^{-d/2-\varepsilon}) \quad \text{for some } \varepsilon > 0. \tag{4.53}$$

Chebyshev's inequality and (4.53) imply that

$$\mathrm{P}(||\Delta_t|| \geq \delta_t) = o(t^{-d/2}) \quad \text{where } \delta_t = t^{-\varepsilon/2},$$

and so (4.41) in Assumption A3 is satisfied. It is clear that these arguments carry through for the m-step chain also for any $m \geq 1$.

It is easily seen that

$$\mathrm{E}|\Theta_t \cdot 1| = o(t^{1/2}),$$

and so (4.42) in Assumption A3 is satisfied for any m-step chain.

Verifying Assumption A2: Assumption (4.39) is clearly satisfied for any m-step chain.

Recall that we assume, without loss of generality, that the matrix A has the form of the right-hand side of (4.32). That is $\{V_t\}$ is a VAR(1) process with A_{22} in place of A, and the eigenvalues of A_{22} are inside the unit circle. Therefore the Markov process $\{V_t\}$ is ergodic; see Theorem 3. Thus (4.40) holds. Also these arguments hold for any m-step chain.

Verifying Assumption A4: As stated in the text following (4.34), $m_t = [c\ln(t)]$ for some positive constant c. Let $T_t = t - m_t$ and from the text above (4.44), the $d \times d$ matrix $\Gamma_t = T_t^{-1}\mathrm{Cov}(Z_t^\star), t \geq 1$. We easily see that $\lim_{t\to\infty} t^{-1}\mathrm{Cov}(\Theta_t) = 0$. Thus, because $Z_t^\star$ and Θ_t are independent and $T_t \sim t$,

$$\Gamma_t \sim t^{-1}\mathrm{Cov}(Z_t) \quad \text{as } t \to \infty. \tag{4.54}$$

When B_1 defined in (3.2) is diagonal, Lemma 6 gives that

$$\lambda_{i,t} \sim tc_i \quad \text{for } i = 1, \ldots d$$

as $t \to \infty$, where $\lambda_{1,t}, \ldots, \lambda_{d,t}$ are the d largest eigenvalues of $\Sigma_t^X = \mathrm{Cov}(X_t | X_0 = 0)$ and $c_1, \ldots, c_d$ are finite and positive constants. Therefore, and by (4.54), we find that (4.44) of Assumption A4 is satisfied when B_1 is diagonal.

To simplify the notation let $\{\zeta_t; t \geq 0\}$ be a sequence of i.i.d. variables independent of $\{\varepsilon_t; t \geq 1\}$ and let ζ_1 and ε_1 be identically distributed. Then $A^k\zeta_k$ and $A^k\varepsilon_{t-k}$ have the same distribution for $k < t$ and $t \geq 1$. Therefore,

$$Z_t^\star \stackrel{\mathcal{L}}{=} \sum_{k=m_t}^{t-1} \xi_k + \mu_t^Z, \qquad t \geq 1, \tag{4.55}$$

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

where $\stackrel{\mathcal{L}}{=}$ means equal in distribution and ξ_k is the vector consisting of the d first components of $A^k\zeta_k$ for $k \geq 0$. That is, $\xi_k = A^k_{d\times n}\zeta_k$, where $A^k_{d\times n}$ is the $d \times n$ matrix consisting of the d first rows of A^k. Recall that A has the form (4.32). Note that the sum in (4.55) consists of T_t terms, so that (4.55) has the form of (4.43) of Assumption A4.

If (4.45) and (4.46) are not satisfied for $m = 1$, we apply the m-step chain; i.e., we let $\xi_t \stackrel{\text{def}}{=} \xi^m_t = \sum_{j=0}^{m-1} A^{mt-j}_{d\times n}\zeta_{mt-j}$. We use Lemma 8 with $A^{(j)}_t = A^{mt-j}_{d\times n}$. By Lemma 3 there is an m such that $\Sigma^{\xi}_t = \mathrm{Cov}(\xi^m_t)$ is positive definite. Thus there is a finite m such that the conditions in Lemma 8 are satisfied, which in turn gives (4.45) and (4.46). ■

LEMMA 10. *Let $Z_t = Z^\star_t + \Theta_t$, $t \geq 1$, be a process with state space $(R^d, \mathcal{B}^d)$ and let $V_t = \Delta_t + V^\star_t$, $t \geq 1$, be a process on the state space $(R^{n-d}, \mathcal{B}^{n-d})$, $n \geq d \geq 1$, with $Z^\star_t$, Θ_t, Δ_t, and $V^\star_t$, as defined in (4.35). Let $D \in \mathcal{B}^{d^+}$ and $S \in \mathcal{B}^{n-d^+}$ be compact balls. Assume that A4 holds; then*

$$\mathrm{P}_0(Z_t \in D) = \mathrm{P}_0(U_t \in D) + \kappa_t(1 + t^{-1/2}||\mu^Z_t||)^{-3}, \qquad t \to \infty. \tag{4.56}$$

where U_t is $\mathcal{N}(\mu^Z_t, T_t\Gamma_t)$ distributed and $\kappa_t = o(t^{-d/2})$.

Assume that A1–A4 hold. Then,

$$\mathrm{P}_0(Z_t \in D, V_t \in S) = \mathrm{P}_0(Z_t \in D)\mathrm{P}_0(V_t \in S) + o(t^{-d/2}), \qquad t \to \infty, \tag{4.57}$$

and

$$\mathrm{P}_0(Z_t \in D, V_t \in S) = c_{D,S}t^{-d/2} + o(t^{-d/2}), \qquad t \to \infty, \tag{4.58}$$

where $c_{D,S}$ is a positive constant depending on D and S.

Proof. In this proof we will write P and E instead of P_0 and E_0, respectively, to ease the notation.

We start by looking at the process $\{Z_t\}$. Our assumptions about the sequence $\{Z^\star_t; t \geq 1\}$, i.e., (4.43)–(4.46), mean that Corollary 19.4 in Bhattacharya and Rao (1976, p. 197) can be used. Therefore,

$$\sup_{z\in R^d}(1 + ||z - T_t^{-1/2}\mu^Z_t||^3)|q_t(z) - \phi_t(z)| \to 0 \quad \text{as } t \to 0, \tag{4.59}$$

where $q_t(z)$ is the density function of $Z^\star_t/\sqrt{T_t}$ and $\phi_t(z)$ is the normal $\mathcal{N}(T_t^{-1/2}\mu^Z_t, \Gamma_t)$ density function of dimension d.

Let $T = T_t$. In what follows we will, without further comment, make extensive use of the fact that because $T = T_t \sim t$,

$$O(T^p) = O(T^p_t) = O(t^p) \quad \text{and} \quad o(T^p) = o(T^p_t) = o(t^p) \quad \text{for any } p.$$

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

For a ball $D \in \mathcal{B}^d$, define

$$D_T = \{z \in R^d : T^{1/2} z \in D\}.$$

Then $\mathrm{P}(Z_t^\star \in D) = \mathrm{P}(T_t^{-1/2} Z_t^\star \in D_T)$ and

$$\int_{D_T} dz = O(T^{-d/2}) = O(t^{-d/2}). \tag{4.60}$$

Now, by (4.59), there is a sequence $\{\tilde{\delta}_t\}$, where $\lim_{t\to\infty} \tilde{\delta}_t = 0$, such that uniformly in z,

$$\begin{aligned}
\mathrm{P}(Z_t^\star \in D - z) &= \mathrm{P}(T^{-1/2} Z_t^\star \in D_T - T^{-1/2} z) \\
&= \int_{D_T} q_t(s - T^{-1/2} z) ds \\
&= \int_{D_T} [\phi_t(s - T^{-1/2} z) + \tilde{\delta}_t] ds \\
&= \mathrm{P}(U_t \in D_T - T^{-1/2} z) + o(t^{-d/2}),
\end{aligned} \tag{4.61}$$

where $U_t \overset{\mathcal{L}}{=} \mathcal{N}(T_t^{-1/2} \mu_t^z, \Gamma_t)$ and the last equality follows by (4.60).

We can now prove (4.56) from (4.61). Let $z = 0$ in (4.61). Because D is bounded, we can by (4.59) choose the sequence $\{\tilde{\delta}_t\}$ such that $\tilde{\delta}_t = c_t(1 + t^{-1/2}||\mu_t^z||)^{-3}$ where $\lim_{t\to\infty} c_t \to 0$. Therefore (4.56) follows.

Next we look at

$$\begin{aligned}
\mathrm{P}(Z_t \in D, V_t^\star \in S) &= \mathrm{P}(Z_t^\star + \Theta_t \in D, V_t^\star \in S) \\
&= \int \mathrm{P}(Z_t^\star \in D - z) \mathrm{P}(V_t^\star \in S, \Theta_t \in dz) + o(t^{-d/2}) \\
&= \int \mathrm{P}(U_t \in D_T - T^{-1/2} z) \mathrm{P}(V_t^\star \in S, \Theta_t \in dz) \\
&\quad + o(t^{-d/2}),
\end{aligned} \tag{4.62}$$

where the second equality follows by (4.38) and the last equality by (4.61).

Define

$$f_t(z) \overset{\text{def}}{=} \mathrm{P}(U_t \in D_T - T^{-1/2} z) = \int_{D_T} \phi_t(s - T^{-1/2} z) ds, \qquad z \in R^d. \tag{4.63}$$

By the mean value theorem

$$f_t(z) = f_t(0) + \bigtriangledown f_t(c_t z) \cdot z \quad \forall z \in R^d, \tag{4.64}$$

with $||c_t z|| < ||z||$ and $c_t z = c_t(z) \in R^d$; i.e., c_t is a function of z for $t \geq 1$, $\bigtriangledown f_t$ is the gradient of f_t, and $\cdot$ is the euclidean inner product. By (4.44) in Assumption A4, it follows that there are finite constants M and t_1 such that

$$\sup_{z \in R^d} |\bigtriangledown \phi_t(z) \cdot 1| \leq M \quad \text{for } t \geq t_1, \tag{4.65}$$

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

where 1 denotes the constant vector in R^d with one in each direction. Combining (4.60) and (4.65), we find that

$$\sup_{z\in R^d}\int_{D_T} |\bigtriangledown \phi_t(s-z)\cdot 1|ds = O(t^{-d/2}). \tag{4.66}$$

By the chain rule and (4.66), we find that (uniformly in z)

$$|\bigtriangledown f_t(c_t z)\cdot 1| = |T^{-1/2}\int_{D_T} \bigtriangledown\phi_t(s-T^{-1/2}c_t z)\cdot 1\, ds| = O\left(t^{-(t+1)/d}\right). \tag{4.67}$$

Because $\mathrm{P}(V_t^\star \in S, \Theta_t \in dz) \le \mathrm{P}(\Theta_t \in dz)$, we obtain

$$\begin{aligned}\int |\bigtriangledown f_t(c_t z)\cdot z|\,\mathrm{P}(V_t^\star \in S, \Theta_t \in dz)| &\le \int |\bigtriangledown f_t(c_t z)\cdot z|\,\mathrm{P}(\Theta_t \in dz)\\ &\le \sup_{i,z} |\bigtriangledown_i f_t(c_t z)|\,\mathrm{E}|\,\Theta_t\cdot 1|\\ &= o(t^{-d/2}),\end{aligned} \tag{4.68}$$

where $\bigtriangledown_i f_t$ is the ith component of the gradient vector. Indeed, the last equality of (4.68) follows from (4.67) and (4.42). By (4.64) and assumption (4.68), we find that as $t\to\infty$,

$$\int f_t(z)\mathrm{P}(V_t^\star \in S, \Theta_t = z)dz = f_t(0)\mathrm{P}(V_t^\star \in S) + o(t^{-d/2}). \tag{4.69}$$

Hence (4.62), (4.63), and (4.69) give that

$$\mathrm{P}(Z_t \in D, V_t^\star \in S) = \mathrm{P}(U_t \in D_T)\mathrm{P}(V_t^\star \in S) + o(t^{-d/2}). \tag{4.70}$$

As stated after (4.61), $T_t^{1/2}U_t \stackrel{\mathcal{L}}{=} \mathcal{N}(\mu_t^z, T_t\Gamma_t)$, so (4.44) ensures that we can use Lemma 4 to the right-hand side of (4.70). We get that

$$\mathrm{P}(Z_t \in D, V_t^\star \in S) = t^{-d/2}c_D\mathrm{P}(V_t^\star \in S) + o(t^{-d/2}), \tag{4.71}$$

where c_D is a positive constant depending on D.

In particular by choosing $S = R^{n-d}$ in (4.70) and (4.71), we have that

$$\mathrm{P}(Z_t \in D) = \mathrm{P}(U_t \in D_T) + o(t^{-d/2}) = t^{-d/2}c_D + o(t^{-d/2}). \tag{4.72}$$

Now we turn our attention to the processes $\{V_t\}$, $\{V_t^\star\}$, and $\{\Delta_t\}$. Let $\delta > 0$ and define

$$a_\delta \stackrel{\mathrm{def}}{=} \mathrm{E}[1_S(V_t)1_{(||\Delta_t||>\delta)}] \le \mathrm{E}[1_{(||\Delta_t||>\delta)}]. \tag{4.73}$$

Using (4.73) we can write

$$\mathrm{E}[1_S(V_t)1_{(||\Delta_t||\le\delta)}] \le \mathrm{P}(V_t \in S) \le \mathrm{E}[1_S(V_t)1_{(||\Delta_t||\le\delta)}] + a_\delta. \tag{4.74}$$

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

Let d_S be the diameter of the compact ball S. For $\delta > 0$, let S_δ and S^δ be the compact balls with diameter $d_S - \delta$ and $d_S + \delta$, respectively. That is, S_δ is slightly less than S, and S^δ is slightly bigger than S. We have that

$$\{V_t^\star \in S_\delta\} \setminus \{||\Delta_t|| > \delta\} = \{V_t^\star \in S_\delta, ||\Delta_t|| \le \delta\} \subset \{V_t \in S, ||\Delta_t|| \le \delta\} \subset \{V_t^\star \in S^\delta\}. \tag{4.75}$$

So by (4.75) and the assumption (4.41), we have that

$$\mathrm{E}[1_{S_\delta}(V_t^\star)] - b_\delta \le \mathrm{E}[1_S(V_t) 1_{(||\Delta_t|| \le \delta)}] \le \mathrm{E}[1_{S^\delta}(V_t^\star)], \tag{4.76}$$

where

$$b_\delta = P(\{||\Delta_t|| > \delta\}).$$

By applying (4.74) to the mid term of (4.76), we get

$$\mathrm{E}[1_{S_\delta}(V_t^\star)] - b_\delta \le \mathrm{E}[1_S(V_t)] \le \mathrm{E}[1_{S^\delta}(V_t^\star)] + a_\delta. \tag{4.77}$$

Now, let $\{\delta_t; t \ge 1\}$ be a sequence such that $\delta_t = o(1)$ and such that (4.41) is satisfied. Letting $\delta = \delta_t$, (4.41) and (4.73) give that

$$a_{\delta_t} = o(t^{-d/2}) \quad \text{and} \quad b_{\delta_t} = o(t^{-d/2}). \tag{4.78}$$

By adding $Z_t \in D$ in (4.77) and using (4.78), we get

$$\begin{aligned} &\mathrm{E}[1_D(Z_t) 1_{S_\delta}(V_t^\star)] - o(t^{-d/2}) \\ &\quad \le \mathrm{E}[1_D(Z_t) 1_S(V_t)] \le \mathrm{E}[1_D(Z_t) 1_{S^\delta}(V_t^\star)] + o(t^{-d/2}). \end{aligned} \tag{4.79}$$

Now, by applying (4.71) with S_δ and S^δ instead of S, respectively, in (4.79), we have that

$$\begin{aligned} t^{-d/2} c_D \mathrm{P}(V_t^\star \in S_\delta) &\le \mathrm{P}(Z_t \in D, V_t \in S) + o(t^{-d/2}) \\ &\le t^{-d/2} c_D \mathrm{P}(V_t^\star \in S^\delta) + o(t^{-d/2}). \end{aligned} \tag{4.80}$$

Let $\lim_{\delta \to 0} S^\delta \setminus S_\delta = N$. By Assumption A2 $P(V_t^\star \in N) = 0$; i.e.,

$$\lim_{\delta \to 0} \mathrm{P}(V_t^\star \in S_\delta) = \lim_{\delta \to 0} \mathrm{P}(V_t^\star \in S^\delta) = \mathrm{P}(V_t^\star \in S). \tag{4.81}$$

Combining (4.77) and (4.81) we have that

$$\lim_{t \to 0} |\mathrm{P}(V_t^\star \in S) - \mathrm{P}(V_t \in S)| = 0. \tag{4.82}$$

Therefore by (4.80)–(4.82), we get

$$\mathrm{P}(Z_t \in D, V_t \in S) = t^{-d/2} c_D \mathrm{P}(V_t \in S) + o(t^{-d/2}), \tag{4.83}$$

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

and so (4.58) is proved by (4.83) and assumption (4.40). Finally, by (4.72) and (4.83) we get

$$\mathrm{P}(Z_t \in D, V_t \in S) = \mathrm{P}(Z_t \in D)\mathrm{P}(V_t \in S) + o(t^{-d/2}), \tag{4.84}$$

i.e., (4.57) is proved. ■

Proof of Theorem 4 in the Case of Nonnormally Distributed Residuals. The nonnormal case is proved in two steps, according to whether B_1, defined in (3.2), is diagonal or nondiagonal.

Step 1. The nonnormal case; B_1 is diagonal. When B_1 is diagonal, Lemma 9 ensures that we can use Lemma 10. Let $C = D \times S$, where D and S are the compact unit balls of dimension d and $(n-d)$, respectively, and notice that C is a small set because it is compact with positive Lebesgue measure. Lemma 10 gives that when $b \in G_A$,

$$\mathrm{P}_0(X_t \in C) = t^{-d/2} c_{D,S} + o(t^{-d/2}), \tag{4.85}$$

where $c_{D,S}$ is a positive constant. By (4.85) the arguments applied in the proof of Theorem 4 (i)–(v) in the case of normally distributed residuals (see Section 4.1) can be used in the case of nonnormal residuals also when $b \in G_A$.

When $b \notin G_A$, (4.56) in Lemma 10, Lemma 6, and point (iii) of Lemma 4 ensure that $\{X_t\}$ is transient by the same arguments as in the case of normally distributed residuals (see Section 4.1). That is, point (v) is proved and also points (iii) and (iv) when $b \notin G_A$ and B_1 is diagonal.

Step 2. The nonnormal case; B_1 is nondiagonal. It remains to prove Theorem 4 when B_1 is nondiagonal, i.e., to prove parts (iii)–(v) of Theorem 4 when B_1 is nondiagonal. It means that B_1 is an elementary Jordan matrix or contains an elementary Jordan matrix.

Define

$$\sigma_{i,t}^2 = \mathrm{Var}(X_{i,t}), \qquad t \geq 1,$$

where $X_{i,t}$ is the ith component of the vector X_t for $t \geq 1$. By equation (4.29) there is at least one component $X_{1,t}$, say, such that

$$c_1 t^3 \leq \sigma_t^2 \leq c_2 t^3 \quad \text{for } t \geq 1, \tag{4.86}$$

where c_1 and c_2 are positive constants and $\sigma_t^2 = \sigma_{1,t}^2$.

Let $J = [t/2]$ where $[\cdot]$ is the greatest-integer-lesser-than function. Furthermore, let $a_1^{(j)}$ be the first row vector of the matrix A^j, $j \geq 0$. Define

$$W_{t,0} = \sum_{j=0}^{J+h} a_1^{(j)} \cdot \varepsilon_{t-j},$$

where the adjustment, h, is 1 or 0, respectively, as t is odd or even and $\cdot$ is the euclidean inner product. Furthermore define

$$W_{t,j} = a_1^{(J+h+j)} \cdot \varepsilon_{t-J-h-j} \quad \text{for } j = 1, \ldots, J; \tag{4.87}$$

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

i.e.,

$$X_{1,t} = W_{t,0} + \sum_{j=1}^{J} W_{t,j} \quad \text{for } t \geq 1. \tag{4.88}$$

Note that, by the same arguments as for (4.86), there are positive constants c_3 and c_4 such that

$$c_3 k \leq ||a_1^{(k)}|| \leq c_4 k \quad \text{for } k \geq 0. \tag{4.89}$$

Define

$$v_J^2 = J^{-1} \sum_{j=1}^{J} \text{Var}(W_{t,j}).$$

From (4.87) and (4.89) we find that there are positive constants c_5 and c_6 such that

$$c_5 J \leq v_J \leq c_6 J \quad \text{for } J \geq 1. \tag{4.90}$$

Notice by (4.89) and (4.90) that $v_J^{-1} a_1^{(J+h+j)}$ is bounded for $j \leq J$ and $t \geq 1$. The idea is that a density version of the central limit theorem can be applied to the variables $v_J^{-1} W_{t,j}$, for $j = 1, \ldots, J$, as $t \to \infty$. Let

$$g_{t,j}^W(u) = \text{E} \exp[iu(v_J^{-1} W_{t,j})] \quad \text{for } j = 1, \ldots, J, \qquad t \geq 1.$$

Recall that $J = [t/2]$. It is now easily seen from (4.87), (4.89), (4.90), and Conditions C1–C3 that

$$\overline{\lim_{J \to \infty}} \; J^{-1} \sum_{j=1}^{J} \text{E}|v_J^{-1} W_{t,j}|^3 < \infty, \tag{4.91}$$

$$\sup_{t \geq 1} \sup_{j \leq J} \int |g_{t,j}^W(u)| du < \infty \tag{4.92}$$

and that for all positive numbers a

$$\sup_{t \geq 1} \sup_{j \leq J} \sup_{u > |a|} g_{t,j}^W(u) < 1. \tag{4.93}$$

Now, by (4.91)–(4.93) we can apply Theorem 19.3 of Bhattacharya and Rao (1976, p. 194), which ensures that (4.30) holds. Therefore, we can conclude that there is a finite constant M such that

$$\sup_{J \geq 1} \sup_{x \in R} q_J(x) \leq M, \tag{4.94}$$

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

where q_J is the density function of

$$J^{-1/2}v_J^{-1}\sum_{j=1}^{J} W_{t,j} \quad \text{for } J \geq 1.$$

Again recall that $J = [t/2]$. Let $C = [-1, 1]$ and $C_J = [J^{-1/2}v_J^{-1}, J^{-1/2}v_J^{-1}]$. Then by (4.90) there is a constant c such that

$$l(C_J) \leq ct^{-3/2}, \tag{4.95}$$

where l is the Lebesgue measure on R.

Let q_J^0 be the density function for

$$J^{-1/2}v_J^{-1}W_{t,0} \quad \text{for } J \geq 1.$$

By (4.88), convolution, and (4.94) we find that

$$\begin{aligned}
\mathrm{P}_0(X_{1,t} \in C) &= \mathrm{P}_0(J^{-1/2}v_J^{-1}X_{1,t} \in C_J) \\
&= \mathrm{P}(J^{-1/2}v_J^{-1}W_{t,0} + J^{-1/2}v_J^{-1}\sum_{j=1}^{J} W_{t,j} \in C_J) \\
&= \int_{C_J}\int q_J(x-z)q_J^0(z)dzdx \\
&\leq \int_{C_J}\int Mq_J^0(z)dzdx \\
&\leq Ml(C_J).
\end{aligned} \tag{4.96}$$

By (4.96) and (4.95) we find that

$$\sum_{t=1}^{\infty}\mathrm{P}_0(X_t \in \tilde{C}) \leq \sum_{t=1}^{\infty}\mathrm{P}_0(X_{1,t} \in C) \leq cM\sum_{t=1}^{\infty}t^{-3/2} < \infty,$$

where $\tilde{C} \in R^n$ is the ball with radius one. So $\{X_t\}$ is transient by Theorem 1. Note that the proof does not depend on the mean, i.e., the proof holds for any vector b in (3.1). ■

REFERENCES

Aparicio, F.M., A. Escribano, & A.E. Sipols (2006) Range unit-root (RUR) tests: Robust against nonlinearities, error distributions, structural breaks and outliers. *Journal of Time Series Analysis* 27, 545–576.

Bec, F., A. Guay, & E. Guerre (2008) Adaptive consistent unit-root tests based on autoregressive threshold model. *Journal of Econometrics* 127, 94–133.

Bhattacharya, R.N. & R.R. Rao (1976) *Normal Approximation and Asymptotic Expansions.* Wiley.

Bingham, N.H., C.M. Goldie, & J.L. Teugels (1987) *Regular Variation.* Cambridge University Press.

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

Chan, K.S. & H. Tong (1985) On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Advances in Applied Probability* 17, 666–678.

Chen, J., J. Gao, & D. Li (2009) Semiparametric Regression Estimation in Null Recurrent Time Series. Manuscript, Department of Economics, University of Adelaide.

Cline, D.B.H. & H.H. Pu (1998) Verifying irreducibility and continuity of a nonlinear time series. *Statistics and Probability Letters* 40, 139–148.

Cline, D.B.H. & H.H. Pu (1999) Stability of nonlinear AR(1) time series with delay. *Stochastic Processes and Their Applications* 82, 307–333.

Davidson, J. (2009) When is a time series I(0)? In J. Castle and N. Shephard (eds.), *A Festschrift for David Hendry,* pp. 322–342. Oxford University Press.

Escribano, A. (2004) Nonlinear error correction: The case of money demand in the United Kingdom (1878–2000). *Macroeconomic Dynamics* 8, 76–116.

Escribano, A., A.E. Sipols, & F.M. Aparicio (2006) Nonlinear cointegration and nonlinear error correction: Record counting cointegration tests. *Communications in Statistics—Simulation and Computation* 35, 939–956.

Feigin, P.D. & R.L. Tweedie (1985) Random coefficient autoregressive processes: A Markov chain analysis of stationarity and finiteness of moments. *Journal of Time Series Analysis* 6, 1–14.

Gao, J., M. King, Z. Lu, & D. Tjøstheim (2009) Specification testing in nonlinear time series with nonstationarity. *Annals of Statistics* 37, 3893–3928.

Granger, C.W.J. (1995) Modelling nonlinear relationships between extended-memory variables. *Econometrica* 63, 265–279.

Granger, C.W.J. & J. Hallman (1991a) Long memory series with attractors. *Oxford Bulletin of Economics and Statistics* 53, 11–26.

Granger, C.W.J. & J. Hallman (1991b) Nonlinear transformations of integrated time series. *Journal of Time Series Analysis* 12, 207–224.

Granger, C.W.J. & N. Swanson (1996) Further developments in the study of cointegrated variables. *Oxford Bulletin of Economics and Statistics* 58, 537–553.

Hamilton, J.D. (1994) *Time Series Analysis.* Princeton University Press.

Horn, R.A. & C.R. Johnson (1990) *Matrix Analysis.* Cambridge University Press.

Juhl, T. & Z. Xiao (2005a) Partially linear models with unit roots. *Econometric Theory* 21, 877–906.

Juhl, T. & Z. Xiao (2005b) Testing for cointegration using partially linear models. *Journal of Econometrics* 124, 363–394.

Kallianpur, G. & H. Robbins (1954) The sequence of sums of independent random variables. *Duke Mathematical Journal* 21, 285–307.

Karlsen, H.A., T. Myklebust, & D. Tjøstheim (2007) Nonparametric estimation in a nonlinear cointegration type model. *Annals of Statistics* 35, 252–299. Earlier version (2000) available as a report under Sonderforschungsbereich 373, Humboldt-Universitat zu Berlin.

Karlsen, H.A., T. Myklebust, & D. Tjøstheim (2010) Nonparametric regression estimation in a null recurrent time series. *Journal of Statistical Planning and Inference* 140, 3619–3626.

Karlsen, H.A. & D. Tjøstheim (2001) Nonparametric estimation in null recurrent time series. *Annals of Statistics* 29, 372–416. Earlier version (1998) available as a report under Sonderforschungsbereich 373, Humboldt-Universitat zu Berlin.

Meyn, S.P. & R.L. Tweedie (1993) *Markov Chains and Stochastic Stability.* Springer-Verlag.

Nummelin, E. (1984) *General Irreducible Markov Chains and Non-negative Operators.* Cambridge University Press.

Park, J.Y. & P.C.B. Phillips (1999) Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory* 15, 269–298.

Park, J.Y. & P.C.B. Phillips (2001) Nonlinear regressions with integrated time series. *Econometrica* 1, 117–161.

Phillips, P.C.B. (1987) Time series regression with a unit root. *Econometrica* 55, 277–301.

Saikkonen, P. (2005) Stability results for nonlinear error correction models. *Journal of Econometrics* 127, 69–81.

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press

Saikkonen, P. (2007) Stability of mixtures of vector autoregressions with autoregressive conditional heteroskedasticity. *Statistica Sinica* 17, 221–239.

Saikkonen, P. & I. Choi (2004) Cointegrating smooth transition regression. *Econometric Theory* 20, 301–340.

Sancetta, A. (2009) Nearest neigbor conditional estimation for Harris recurrent Markov chains. *Journal of Multivariate Analysis* 100, 2224–2236.

Schienle, M. (2007) Reaching for Econometric Generality: Nonparametric Nonstationary regression. Working paper, Department of Economics, University of Mannheim.

Spitzer, F. (1964) *Principles of Random Walk.* Van Nostrand.

Stewart, G.W. & J.-G. Sun (1990) *Matrix Perturbation Theory.* Academic Press.

Tjøstheim, D. (1990) Nonlinear time series and Markov chains. *Advances in Applied Probability* 22, 587–611.

Tweedie, R.L. (1976) Criteria for classifying general Markov chains. *Advances in Applied Probability* 8, 737–771.

Wang, Q.Y. & P.C.B. Phillips (2009) Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* 25, 710–738.

https://doi.org/10.1017/S0266466611000119 Published online by Cambridge University Press