Representation Theory on Vector Error Correction Models with Nonlinear Adjustments (DRAFT)*

M. Rodrigo Dupleich-Ulloa†
Department of Economics, University of Warwick

December 2003

Abstract

This paper studies the partial extensions to the Granger Representation Theorem for nonlinear cointegrated systems. The analysis focuses on Vector Error Correction models with nonlinear adjustments using both Markov chain theory, similar to Bec and Rahbek (2002), and Near Epoch Dependence processes, similar Escribano and Mira (2002). The paper shows that these two strands in the literature of asymptotic theory on dynamic nonlinear models are closely linked, particularly on stochastic stability conditions (e.g. Lyapunov exponents).

*I would like to thank Jeremy Smith for constructive comments. Obviously, any errors remains mine.
†Address: Department of Economics, University of Warwick, Coventry-UK, CV4 7AL. email: M.R.Dupleich-Ulloa@warwick.ac.uk
1 Introduction

Economic theory has suggested some potential sources for the presence of nonlinearities in economic systems, such as heterogeneity of economic agents, multiple equilibrium, self-fulfilling expectations, hysteresis, endowment regime shifts, etc. However, in empirical economic analysis, many of these theories are analyzed in linear statistical framework. Therefore, nonlinear time series methods have been trying to fill the gap between nonlinear economic theory and linear empirical evidence.

In a linear environment, comovement among nonstationary variables is called cointegration. This concept introduced by Granger (1981) and Engle and Granger (1987), attempts to establish long-run statistical equilibrium relationships across nonstationary variables. Although, this measure of comovement in multivariate systems has typically been assumed to be linear, there exist theoretical reasons and empirical evidence that in some cases the structure of this comovements is inherently nonlinear. Therefore, this paper studies the partial extensions to the Granger Representation Theorem for nonlinear cointegrated systems.

The study of these nonlinear phenomena for nonstationary processes has started to develop rapidly in the econometrics literature, and has been mainly application-oriented. A possible explanation lies in the difficulties in building asymptotic theory of nonlinear transformations and linking the notions of nonstationarity and nonlinearity. As shown in Granger (1995), the concept of stationarity in terms of integrated processes is not sufficient to handle stochastic stability analysis in a nonlinear framework.

As a result, the first part of the paper develops a partial extension of Granger representation theorem within a Markov chain environment, similar to Bec and Rahbek (2002) and Saikonnen (2001). Using standard Markov Chain theory, we analyze the stochastic properties of a nonlinear cointegrated system by examining the stochastic stability conditions an ergodicity and stationarity are obtained in a basic vector error correction (VEC) model with nonlinear adjustments, assuming that the cointegrated vectors are known.

As pointed out by Escribano and Mira (2000), stationary ergodicity can not realistically be determined and this assumption is relaxed in the second part of the by allowing the data generating process to exhibit not only temporal dependence but also certain forms of temporal heterogeneity. As a result, concepts of mixing conditions and near epoch dependence (NED) are employed to specify more primitive disturbances so that the stochastic process has a “fading” memory.

Similar to Escribano and Mira (2002), we partially extend Granger representation theorem using the concept of Near epoch dependence (NED) on mixing
processes, which constitutes the second approach to examine asymptotic theory in dynamic nonlinear models. Furthermore, by making the process NED a function of a mixing process, serial dependence of the process is restricted sufficiently, so the process is a mixingale, an asymptotic analogue of a martingale. Therefore, this result makes it straightforward to establish (Functional) Central Limit theorems (CLT) and (Uniform) Law of Large Numbers (LLN).

The paper shows that these two strands in the literature of asymptotic theory on dynamic nonlinear models are closely linked, particularly on stochastic stability conditions. The results based on Markov chain theory are more restrictive and nest the ones obtained with NED processes approach, as showed in the statistical theory literature. As the class of nonlinear time-series models and methods is a very large field, we concentrate our discussion on the class of smooth transition (STAR), threshold (TAR) and autoregressive conditional root (ACR) processes.

The plan of this paper is as follows. In section 2, we briefly analyze the basic result on linear cointegration. Section 3 is devoted to present a partial extension of Granger representation theorem using Markov chain theory and NED. Section 4 contains conclusions and research avenues for future extensions.

2 Linear Cointegration

To understand the idea behind linear cointegration, we define $X_t$ as a $p$-vector of $I(1)$ time series in which linear combinations (i.e. $\beta'X_t$) appear to be stationary. Those variables are said to be cointegrated and the weight ($\beta$) is the cointegrating vector. When the variables are cointegrated, they share common stochastic trends that drive their long swings and the linear combinations annihilate the common stochastic trends.

Consider a $p$-dimensional VAR($k$) process for $X_t$, this can be written in a error correction form,

$$\Delta X_t = \alpha\beta'X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t$$

(1)

where $t = 1, 2, \ldots, T$ and $\alpha$ and $\beta$ are $p \times r$ matrices while $(\Gamma_i)_{i=1,\ldots,k-1}$ are $p \times p$ matrices, where $\alpha\beta' = \sum_{i=1}^{k} A_i - I_p$, $\Gamma_i = -\sum_{j=i+1}^{k} A_j$ and $(A_i)_{i=1,\ldots,k}$ are $p \times p$ matrices of the underlying VAR($k$) model. The sequence $\varepsilon_t$ is assumed to be i.i.d. $(0, \Omega)$ with $\Omega > 0$ and $(X_{-i})_{i=0,\ldots,k-1}$ are fixed.

The main stability and stationarity properties are addressed in Granger representation theorem. Assume the rank of $\alpha$ and $\beta$ equals $r < p$ and the characteristic polynomial,
\[ A(z) = I_p (1 - z) - \alpha \beta' z - \sum_{i=1}^{k-1} \Gamma_i (1 - z) z^i \]

has exactly \( p - r \) roots at \( z = 1 \), while the remaining roots are larger than one in absolute value, \(|z| > 1\), or equivalently, that,

\[ |\alpha' \Gamma \beta'| \neq 0 \]

where, for example, \( \alpha \perp \) is an orthogonal complement of \( \alpha \) such that, \( \alpha' \alpha \perp = 0 \). Then it can be proved that \( X_t \) is a I(1) cointegrated process and \( \Delta X_t \) and \( \beta' X_t \) are stationary given an initial distribution. Furthermore, the solution \( X_t \), to (1) has the representation,

\[ X_t = C \sum_{i=1}^{t} \varepsilon_i + u_t + A \]

Here the long run impact matrix is \( C = \beta \perp (\alpha' \Gamma \beta \perp)^{-1} \alpha' \perp, u_t = \sum_{i=0}^{\infty} C_i \varepsilon_{t-i} \) is a stationary linear process with exponentially decreasing coefficients and \( A \) depends on initial conditions, and satisfies \( \beta' A = 0 \) (details are given in Johansen (1996)). In this framework, \( X_t \) satisfies FCLT and therefore, asymptotic theory for inference on parameters and cointegrating rank can be expressed in terms of Brownian motions.

Furthermore, the common trends in (3) are the variables \( \alpha' \perp \sum_{i=1}^{t} \varepsilon_i \) and the linear combination \( \alpha' \perp \Delta X_t = \alpha' \perp \varepsilon_t \) annihilate the common cycles proposed by Vahid and Engle (1993).

### 3 Nonlinear Cointegration

Similar to Bec and Rahbek (2002), a nonlinear error correction (NEC) model can be formulated as follows,

\[ \Delta X_t = F (\beta' X_t) + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t \]

where \( t = 1, 2, \ldots, T \), \( \beta \) is a \( p \times r \) matrix and \( (\Gamma_i)_{i=1, \ldots, k-1} \) are \( p \times p \) matrices, similar to the linear case. \( F : \mathbb{R}^r \to \mathbb{R}^p \) is a nonlinear mapping of \( \beta' X_t \) and the sequence zero mean disturbances \( \varepsilon_t \) with covariance matrix \( \Omega > 0 \) (detailed assumptions are specified later). As in the linear case, the initial values \( (X_{-1})_{i=0, \ldots, k-1} \) are fixed. Clearly, the model in (4) nests the linear I(1) model in (1), if,

\[ F (\beta' X_t) = \alpha \beta' X_{t-1} \]
A more general form of the model in conditional mean given in (4), is given by,

$$\Delta X_t = f \left( \beta' X_{t-1}, (\Delta X_{t-i})_{i=1,\ldots,k-1} \right) + \varepsilon_t$$

(5)

The process above can be seen as a nonlinear process with additive noise as discussed in Tong (1990). Similar to Tjøstheim (1990), it can proven using Markov chain theory that the equation above is governed by a geometric ergodic process. A stability condition is required such that \( f(\cdot) \) is a contraction mapping (i.e. \( \partial f(\cdot)/\partial X < 1 \)). Analogously, as showed in Davidson (1994) and Gallant and White (1988), using the identical stability condition, the process given in (5) has a “fading memory” in a “near epoch dependence” sense. These ideas are introduced and discussed in the subsequent sections. However, it should be pointed out that an aim of the paper is to show that the two strands in the literature of asymptotic theory in dynamic nonlinear models are linked through stability conditions.

In this paper, we give a specific functional form to the nonlinear map, such that,

$$f \left( \beta' X_{t-1}, (\Delta X_{t-i})_{i=1,\ldots,k-1} \right) =$$

(6)

\[ s_t \left[ \alpha \beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} \right] + (1 - s_t) \left[ a \beta' X_{t-1} + \sum_{i=1}^{k-1} G_i \Delta X_{t-i} \right] + \varepsilon_t + s_t \left( a + s_t(\alpha - a) \right) \beta' X_{t-1} + \sum_{i=1}^{k-1} (G_i + s_t(G_i - G_i)) \Delta X_{t-i} + \varepsilon_t \]

where \( s_t \) is a zero-one valued transition function. For the ACR model, \( s_t \) is specified as follows,

$$P \left( s_t = 1 \mid \beta' X_{t-1}, \varepsilon_t \right) = p \left( \beta' X_{t-1} \right)$$

for the TAR case,

$$s_t = 1 \ (\beta' X_{t-1} > c)$$

and the STAR case is given by,

$$s_t = G \left( \beta' X_{t-1}; \gamma^*, c \right)$$

(7)

Define \( \| \cdot \| \) such that if \( x \) is a \( p \)-dimensional vector, then \( \| x \| \) denotes a vector norm, e.g. a Euclidean norm such that \( \| x \| = x'x \). Likewise, for a matrix \( X \), a matrix norm is expressed \( \| X \| \), e.g. in a Euclidean space \( \| x \| = tr \left( X'X \right) \). Bek and Rahbek (2002) allow the regimes switches to be determined by \( \| \beta' X_{t-1} \| \), the squared sum of the \( r \) disequilibrium errors (i.e. euclidean distance to the attractor set).

In our standard VEC with nonlinear adjustments (equation (5) and (6)), it is assumed that the dynamics of the system can be approximated by a 2-piecewise linear approximation. More simply, the system switches between two
regimes or states of the world as $s_t$ travels from zero to one. A multiple regime threshold vector error correction model (TVECM) is analyzed in Tsay (1998), De Gooijer and Vidiella-i-Anguera (2002), Lo and Zivot (2001). However, we initially analyze a nonlinear representation theory with a standard 2-regime model in a Markov chain framework, followed by the extension proposed by Escribano and Mira (2002) using near epoch dependent processes (NED).

3.1 Representation Theory based on Markov Chain Theory

In this section, some basic definitions and tools of Markov Chain theory are introduced to analyze the stochastic properties of a nonlinear dynamic system, such as stability and stationarity. The analysis implies that stochastic stability, under certain conditions, will ensure stationarity and ergodicity of a stochastic process. As a result Central Limit Theorems (CLT) and Law of Large Numbers (LLN) can be employed to derive asymptotic normality, consistency of estimators and inference of the model. Moreover, these tools will be important later to generalize the idea of cointegration. This section is based on Rahbek and Shephard (2001), Tjøstheim (1990), Chan and Tong (1985), and an extensive reference to this theory are treated in Meyn and Tweedie (1993), Nummelin (1984) and and Appendix 1 of Tong (1990).

3.1.1 Elements of Markov chain theory

Some basic tools of Markov chains theory are summarized in this subsection, defined on a general measurable space (e.g. Euclidean). Let $(X_t, t \geq 0)$ be a time homogeneous Markov chain with state space $(\mathbb{R}^p, B^p)$, where $B^p$ is the Borel $\sigma$–algebra on $\mathbb{R}^p$, and define the probability measure $P$ on $(\mathbb{R}^p, B^p)$. The $k$th step transition probability for $k \geq 1$ is denoted by $P^k(x, A)$, that is,

$$P^k(x, A) = P(X_k \in A \mid X_0 = x) = P(X_{m+k} \in A \mid X_m = x)$$

where $x \in \mathbb{R}^p$ and $A \in B^p$ for all $m \geq 0$. The condition above implies that the stochastic process $(X_t, t \geq 0)$ has the Markov property; that is, at every $m$, the next state $X_{m+k}$ depends only on the present state $X_m$ of the process. The following regularity condition is required to build our analysis,

**Condition 1.** For some $k \geq 1$, the $k$th step transition probability has a strictly positive and continuous density with respect to a Lebesgue measure $\phi (\mathbb{R}^p) > 0$ on $(\mathbb{R}^p, B^p)$, i.e.

$$P^k(x, A) = P(X_{m+k} \in A \mid X_m = x) = \int_A f(y \mid x) \, dy > 0,$$
for all \( x \in \mathbb{R}^p, A \in B^p \) and \( m \geq 0 \)

The next lemma makes use of Condition 1, so a Markov chain is irreducible with respect to the Lebesgue measure, aperiodic and compact sets \( C \subset \mathbb{R}^p \) are small. Similar presentation is given in propositions A1.1 to A1.4 in Tong (1990).

**Lemma 1.** A time homogeneous Markov chain \((X_t, t \geq 0)\) on \((\mathbb{R}^p, B^p)\) that satisfies Condition 1 for some \( k \geq 1 \) is \( \phi \)-irreducible, aperiodic and compact sets \( C \subset \mathbb{R}^p \) are small.

The proof can be found in Nummelin (1984), Meyn and Tweedie (1993) and Rahbek and Shephard (2002). However, a compact set \( C \) is “small” if for some \( m = m(A) \geq 1, \inf_{x \in C} \sum_{k=1}^{\infty} P^k(x, A) > 0, \) for all \( C \subset \mathbb{R}^p \) and \( A \in B, \) with \( \phi(A) > 0. \) This property is needed since our final objective is to set regularity conditions under which the chain will return to these “small” compact sets.

Irreducibility with respect a Lebesgue measure \( \phi, \) guarantees that almost all “reasonable sized” sets (as measured by \( \phi \)) are always reached by the chain with some positive probability, from every possible starting point, therefore, the chain cannot be divided into separate “reduced” pieces. Formally, \((X_t, t \geq 0)\) is \( \phi \)-irreducible, if \( \sum_{k=1}^{\infty} P^k(x, A) > 0, \) for all \( x \in \mathbb{R}^p \) and \( A \in B^p, \) with \( \phi(A) > 0. \)

Aperiodicity in a \( \phi \)-irreducible Markov chain implies that a state space \( \mathbb{R}^p \) that cannot be divided into disjoints sets which the chain jumps results in an aperiodic Markov process. Roughly speaking, we are preventing the possibility that the chain returns to given states only at specific time points. Given a small set \( C \) and a positive integer \( m, \) if \( P^m(x, C) > 0 \) and \( P^{k+1}(x, C) > 0, \) for \( x \in C, \) then \((X_t, t \geq 0)\) is said to be an aperiodic Markov chain. Further technical details are develop in Meyn and Tweedie (1993) and Nummelin (1984). As pointed out by Chan and Tong (1985), these conditions will ensure ergodicity on the irreducible chain \((X_t, t \geq 0),\) by identifying a “mean drift” towards a “centre” (asymptotic solution) of the process.

To understand the concept of ergodicity, Meyn and Tweedie (1993) provides a functional analysis regarding the “broad” concept of stochastic stability in term of Markov chains. The first level of stability, defined above, requires \( \phi \)-irreducibility in the chain, implying that for every possible starting point, all “reasonable sized” sets, measured by \( \phi, \) can be reached. That is, in terms of hitting times, \(^1\) require that for every starting point \( x \in \mathbb{R}^p, \) \( \text{P}(\tau_A < \infty \mid X_0 = x) > 0 \) with \( \phi(A) > 0. \)

Roughly speaking, we can think that the assumption of irreducibility prevents the process for sensitive dependence on initial conditions (SDIC), hence

\(^1\text{Define the first time a chain reaches the set, as a hitting time in which } \tau_A = \inf \{ t \geq 1 : X_t \in A \}.\)
the system is characterized by stable starting points. Specifically, given small changes in the initial position of the chain, it is possible to reach the same set of states.

A stronger concept of stability will not only secure that all possible states can be reached from distinct starting points, but in addition the likelihood of reaching these states is assured. This situation is defined as recurrence in the literature of Markov chains, and involves that for every starting point \( x \in \mathbb{R}^p \) and \( \phi(A) > 0 \), \( P(\tau_A < \infty \mid X_0 = x) = 1 \) or \( E(\tau_A \mid X_0 = x) < \infty \) almost surely (a.s.). Under these conditions, the “centre” of the space attracts the chain (i.e. the process does not drift off), under the former condition the recurrence of the process is faster than when the first condition holds. More easily, in a recurrent Markov process, “reasonable sized” sets are reached with probability of one. If the chain is not recurrent, it is said to be transient. See Chapter 8 in Meyn and Tweedie (1993) for a extended discussion.

A stronger level of stability, based on Meyn and Tweedie (1993), requires a convergence of the distributions of a strong recurrent chain. The limiting behavior is governed by a “invariant regime” denoted as a measure \( \pi \); then, \((X_t, t \geq 0)\) is said to be an ergodic chain. That is, if the initial position \( X_0 = x \) is drawn from the distribution \( \pi \), then the chain remains in the “centre” of the space. Contrary, if the chain starts in a different position, the process converges to the “centre” of the space in a strong probabilistic sense with a limiting distribution given by \( \pi \). More formally,

**Definition 1.** \((X_t, t \geq 0)\) is said to be geometric ergodic if,

\[
\|P^k(x, \cdot) - \pi(\cdot)\| \leq \rho^k h(x), \text{ for } x \in \mathbb{R}^p
\]

where \( \pi \) is a probability measure on \((\mathbb{R}^p, B^p)\), a constant \( 0 < \rho < 1 \) and a \( \pi \)-integrable non-negative measurable function \( h \). Moreover, \( \pi(\cdot) \) is an “invariant measure” if the so-called invariant equation is satisfied, such that \( \pi(A) = \int P(x, A) \pi(dx) \), for \( A \in B^p \). Furthermore, if the chain \((X_t, t \geq 0)\) is started with a initial distribution \( \pi \), the process is strict stationary\(^2\). This powerful result helps us to develop asymptotic results, since convergence of distribution is ensured for the chain and (bounded) functions of the chain, as a consequence, LLN and CLT can be applied.

In order to practically establish a criteria that guarantees geometrically ergodicity in a Markov chain, Tweedie (1975) has extended the so-called “Foster-Lyapunov” criteria. In an influential paper Chan and Tong (1985) link the analysis of stability theory of deterministic difference equations with ergodicity of stochastic systems via the concept of a Lyapunov function. The application of this criteria in the literature has been a key feature for identifying the stochastic

\(^2\)Otherwise, the result implies that \((X_t, t \geq 0)\) is asymptotically stationary at an exponential rate.
stability conditions for nonlinear dynamic models, without having to solve the difference equation. Moreover, the criteria was originally motivated by energy considerations, hence a Lyapunov function is sometimes known as a generalized energy.

In this paper, a similar criteria labelled drift criteria for geometric ergodic, developed by Tweedie (1975), is employed for the analysis of stochastic stability. Tjøstheim (1990) explains the advantages of using the later criteria compared to the Lyapunov functions developed in Chan and Tong’s seminal paper.

**Definition 2.** A drift function $V : \mathbb{R}^p \to [1, \infty]$ satisfies a $m$-step geometrical drift criterion (relative to a Markov Chain) if there exists a compact set $C \subset \mathbb{R}^p$, and constants $\beta \in (0, 1)$ and $M < \infty$, such that,

$$E^mV(x) := E(V(X_{t+m}) \mid X_t = x) \leq \begin{cases} \beta V(x), & \text{if } x \in C^c \\ M, & \text{if } x \in C \end{cases}, \text{ for all } x \in \mathbb{R}^p$$

where $V(x)$ is interpreted as a generalized energy function and the compact set $C$ as the centre of attraction. Intuitively, if the Markov chain starts outside $C$ (i.e. $C^c$), the expected value of $V(X_{t+m})$ will be less than $V(x)$, in physics terms the chain will, on average, dissipate energy in the next state and the process is forced to return to the centre of attraction $C$. Alternatively, if initial position of the chain is inside $C$, the average gain in energy in the next state is bounded by $M$.

Similar to Theorem A1.5 in Tong (1990), Theorem 15.0.1 in Meyn and Tweedie (1993) and Theorem 3 in Rahbek and Shepard (2001), the following theorem is an important consequence of the $m$-step geometric drift criteria which will be central for our purposes,

**Theorem 1.** Let $(X_t, t \geq 0)$ be a time homogeneous Markov chain on $(\mathbb{R}^p, B^p)$ which is $\mu$-irreducible, aperiodic and for which compact sets $C \subset \mathbb{R}^p$ are small. Moreover, if a $m$-step geometric drift criteria is satisfied for some drift function $V$, then, the process is geometrically ergodic and there exists an invariant measure for the process. If $X_t$ is initiated at the invariant distribution $\pi$ then the Markov chain is strict stationary and ergodic and $EV(X_t) < \infty$.

The proof to this theorem can be found in Chapter 15 of Meyn and Tweedie (1993), Dennis, Hansen and Rahbek (2002), among others. Two remarks should be made, firstly, the last condition implies that $V(\cdot)$ is integrable, as a result any moment of $X_t$ are bounded by $V(\cdot)$ (see Tjøstheim (1990) for details). Secondly, geometrically ergodic processes are strong mixing, so the present seems to be stronger that the one discussed in second approach.
### 3.1.2 Representation Theory under geometric ergodicity

The analysis is based initially on Markov Chain theory for the model specified in (6) and (7) with \( k = 2 \), such that,

\[
\Delta X_t = s_t \left[ a^* X_{t-1} + \Gamma_1 \Delta X_{t-1} \right] + (1 - s_t) \left[ a^* X_{t-1} + G_1 \Delta X_{t-1} \right] + \varepsilon_t \tag{9}
\]

the sequence of disturbances \( \varepsilon_t \) are assumed to have a zero mean and a positive definite covariance matrix denoted as \( \Omega > 0 \). The strategy to partially extend the Granger representation theorem using Markov chain theory has some similarities to one used in Johansen (1996) in the linear case. That is, we have to prove that \( \Delta X_t \) and \( \beta^* X_{t-1} \) are geometric ergodic and strict stationary chains, hence our stability analysis, similar to Bec and Rahbek (2002), should employ the tools introduced above.

Similar to Johansen (1996), define \( V_t = (X_t^*, \Delta X_t^{\beta_1})' = (Z^\beta', Z_t^{\beta_1'})' \), this turns out to be useful in the general case for generating a homogeneous Markov chain. Therefore, premultiplying (9) by \((\beta, \beta_1)'\), yields,

\[
V_t = s_t (\Phi_1 V_{t-1} + \Phi_2 V_{t-2}^*) + (1 - s_t) (\Phi_1 V_{t-1} + \Phi_2 V_{t-2}) + \eta_t \tag{10}
\]

where \( \overline{\beta} = \beta (\beta^\prime \beta)^{-1} \) and \( \beta^\prime + \beta_1^\prime = I_p \). Note that \( \beta \) is a \( p \times r \) matrix of rank \( r < p \), \( \beta_1 \) denotes a orthogonal complement matrix (i.e. \( \beta_1^\prime = 0 \)) with dimension \( p \times (p - r) \) and rank \( p - r \). Furthermore, \( \eta_t = (\beta, \beta_1)' \varepsilon_t \), where \( \varepsilon_t \sim \text{i.i.d.} (0, \Sigma) \), \( \Sigma = (\beta, \beta_1)' \Omega (\beta, \beta_1) > 0 \) (Details are given in Appendix 1). Equation (10) has to be rewritten in a companion form, in order to obtain a homogeneous Markov chain on \( V_t = (V_t', V_{t-1}')' \). Thus, (10) can be written as follows,

\[
V_t = s_t \Phi_1 V_{t-1} + (1 - s_t) \Phi_2 V_{t-1} + \eta_t^* \tag{11}
\]

where, \( \Phi_1 = \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ I_p & 0 \end{pmatrix} \), \( \Phi_2 = \begin{pmatrix} \Phi_{21} & \Phi_{22} \\ I_p & 0 \end{pmatrix} \), \( \eta_t^* = (I_p, 0)' \eta \) and covariance matrix is given by \( \Sigma^* = \beta^\prime \Omega \beta \). Here, the coefficient matrices \( \Phi_1 \) and \( \Phi_2 \) have dimension of \( p \times 2p \). Simultaneously, the support of the transition function \( s: \mathbb{R}^r \rightarrow [0, 1] \) has to be changed from \( \beta^\prime X_t \) to \( V_t \), therefore a matrix \( \gamma = (I_p, 0, 0, 0)' \) can be defined so \( s_t = p(\gamma' V_{t-1}) \).

In order to apply Lemma 1 on (11), we introduce the following assumptions,

**Assumption 1.** Suppose that,

(a) The stochastic transition variable \( s_t \) is zero-one valued, with switching probability or transition function given by,

\[
P (s_t = 1 \mid V_{t-1}, \eta_t^*) = p(\gamma' V_{t-1})
\]

with \( p(v) \rightarrow 1 \) as \( v \rightarrow \infty \) and \( p(v) \rightarrow 0 \) as \( v \rightarrow -\infty \) and where \( p: \mathbb{R}^{2p} \rightarrow [0, 1] \).
(b) The $\eta_t^*$ sequence is i.i.d. $(0, \Sigma^*), \Sigma^* > 0$ and $E|\eta_t^*|^{2n} < \infty$ for some interger $n \geq 1$. Moreover, $\eta_t^*$ has a positive and continuous density with respect to a Lebesgue measure $\phi (R^{2p}) > 0$ on $(R^{2p}, B^{2p})$.

Under the above assumption, Lemma 1 can be directly applied to (11), as a result the subsequent result holds,

**Proposition 1.** The process on $(V_t, t \geq 0)$ is a time homogeneous Markov chain on $(R^{2p}, B^{2p})$. Moreover, for some $k \geq 1$, the chain is $\phi$-irreducible, aperiodic and compact sets $C \subset R^{2p}$ are small.

The proof is given in appendix 1 and it is based on Chan and Tong (1985, 1986) and Shephard and Rahbek (2002). Through this result, the Markov chain in (11) is stochastic stable in a “weak” sense, consequently all “reasonable sized” sets on $R^{2p}$ can be reached, independently of the initial conditions of the chain.

Based on proposition 1, the next stage has to combine these results with a drift criteria, hence the Markov chain $(V_t, t \geq 0)$ is geometric ergodic process. Therefore, the subsequent assumptions need to be inserted,

**Assumption 2.** For the system given in (11), suppose that,

(a) The autoregressesive function is discontinuous, i.e. there exist a $V^*$ such that,

$$(\Phi_1 - \Phi_2) V^* \neq 0$$ (12)

(b) The maximum absolute value of the eigenvalue of $\Phi_i$ is smaller than one (or spectral radius of $\Phi_i$),

$$\max_{i=1,2} \rho (\Phi_i) < 1$$

or equivalently, the roots of the polynomial, $|I - \Phi_i z - \Phi_i^2 z^2| = 0$ lie outside the unit circle.

In Bec and Rahbek (2002), the regime switching process is governed by $\|\beta X_{t-1}\|$, implying that $p(v) \rightarrow 1$ when the there is a significant distance (euclidean sense) from the long run attractor, i.e. the deviation could be either positive or negative. As a result Assumption 3(ii) in Bec and Rahbek (2002) simply controls for the stability of the outer regime (i.e. when $s_t \rightarrow 1$), the polynomial in the $\Phi_2$-regime is permitted to have either unit or explosive roots.

A more general form is being specified here, where the transition function could map from $s : R^{(r+k_0)} \rightarrow [0, 1]$ depending on $x_t = (X_{t-1}^T, (\Delta X_{t-1})_{i=1,\ldots,k_1})$.  

11
For example, if the deviations to the long run equilibria are significant and positive in all coordinates, \( p(v) \rightarrow 1 \) as \( v \rightarrow \infty \), whereas \( p(v) \rightarrow 0 \) as \( v \rightarrow -\infty \), when the system goes considerably away from equilibria with negative coordinates. As a result, similar to Lemma 3.1 of Chan and Tong (1985) and Proposition 2.1. in of Chan and Tong (1986), Assumption (2b) needs to control for the highest absolute value of the maximal eigenvalue in the polynomial \( |I_p - \Phi_1 z - \Phi_1 z^2| = 0 \) for \( i = 1, 2 \). Moreover, conditional independency between the innovation in \( \eta_t \) and the transition function \( s_t \) is specified in the first part of Assumption (1a).

As it will be showed later, the crucial difference between this approach and the one developed in next section, relies on Assumption (1b), in which the sequence \( \{\eta_t^r\} \) is drawn from an invariant positive and continuous density function, later this assumption is relaxed. Moreover, higher moments up to \( 2n \) where \( n \geq 1 \) exists and leads to the existence of similar higher moments on the chain \( (V_t, t \geq 0) \).

Similar to Chan (1993) and Bec and Rahbek (2002), Assumption (2a) illustrates the nonlinear structure of the stochatic process given in \( (V_t, t \geq 0) \). Based on our results presented in section 3.1.1., the following Proposition holds,

**Proposition 2.** If Assumptions 1 and 2 hold for the Markov chain \( (V_t, t \geq 0) \), then \( (V_t, t \geq 0) \) is a geometrically ergodic process and stationary for a particular initial distribution in \( V_0 \). Moreover, \( E|V_t|^{2n} < \infty \) (i.e. moments up to order \( 2n \) exists).

A sketch of the proof of the theorem is given in appendix 1. However, the requirements of \( \phi \)-irreducibility, aperiodicity and the existence of small compact sets on \( \mathbb{R}^{2p} \) is taken care of by the Assumption (1b) on \( \{\eta_t^r\} \). The proofs are developed in Chan and Tong (1985) and (1986). Bec and Rahbek (2002) have used the following drift criteria,

\[
g(v) = 1 + v'Dv \geq 1, \quad D = \sum_{i=0}^{\infty} \Phi_1^i \Phi_1^i = (I_{2p} - \Phi_1 \Phi_1')^{-1}
\]

For \( k = 2 \), we can specify \( \mu = (0 0 0 I_p)' \) and \( \gamma = (I_r 0 0 0)' \).

This function permits us to obtain the stochastic stability conditions for geometric ergodicity in (9) if the transition variable is defined as \( \|\beta'X_t-1\| \), as a result, we need to control only the roots of the polynomial of the “outer” regime (i.e. \( \rho(\Phi_1) < 1 \)), since \( p(\|v\|) \rightarrow 1 \) when the system deviates significantly from the attractor set, either positively or negatively (see Bec and Rahbek (2002) for details). More formally, given that transition variable is defined as a Euclidean vector norm such that the regime switching process depends on the squared sum of the disequilibrium errors (i.e. the squared distance to the attractor set).

Furthermore, the quadratic form in the drift function ensures the existence of second order moments in \( V_t \) (i.e. \( n = 1 \)). However, if we allow a more
general specification as given in Assumption (1a), the roots of the polynomials of both regimes need to be controlled as specified in Assumption (2b). Thus, the present proof employs the following drift criteria,

\[
g(v) = 1 + \mathbf{1}_{|v'| > 0} v'D_1 v + \mathbf{1}_{|v'| \leq 0} v'D_2 v \geq 1 \tag{13}
\]

\[
D_j = \sum_{i=0}^{\infty} \Phi_j^i \Phi_j^{-1} = (I_{2p} - \Phi_j^j \Phi_j^{-1})^{-1}, \text{ for } j = 1, 2 \tag{14}
\]

This generalized energy function is similar to one analyzed in Chan and Tong (1985), Petruccelli and Woolford (1984) and Tjøstheim (1990) for univariate systems. The proof shows that Definition 2 holds when \( v^2 C > 0 \) (this may imply that \( v^0 D_j v > M_1 \), for \( j = 1, 2 \)), as the expected value of \( g(V_t) \) is smaller than \( g(v) \) at \( V_{t-1} = v \). As a consequence, the process \( (V_t, t \geq 1) \) does not explode over time and a dissipation of energy will take place in the chain at \( C^C \); thus, the process is forced back to return to the attraction set \( C \). Alternatively, on \( C \), \( v^0 D_j v \) is bounded by \( M_1 \) for all \( v^2 < \eta^2 \) and \( j = 1, 2 \), hence \( E[g(V_t) | V_{t-1} = v] \) is continuous and bounded on the compact set \( C \).

Therefore, the Markov chain \( \{V_t\}_{t=1,2,..} \) follows a \( ||v|| \)-geometric ergodic process and if the chain is started with initial distribution \( \pi \), i.e. \( \ell(v_0) = \pi \), then \( (V_t, t \geq 1) \) is strict stationary, based on Theorem 1.

Additionally, similar to Theorem 2.2. in Petruccelli and Woolford (1984), this result implies that the drift function \( E_\pi g(V_t) < \infty \) is integrable with respect to \( v \), therefore the chain \( (V_t, t \geq 1) \) has finite second moments, given the quadratic structure of our drift criteria and Assumption (2a) (i.e. \( E[|\eta|^2] < \infty \) for \( n = 1 \)). As pointed out by Bec and Rahbek (2001), higher moments can be obtained by modifying the structure of the drift criteria function and setting \( n \geq 1 \). The proof is developed in Lemma 6.1 in Tjøstheim (1990), Theorem 2.2. in Petruccelli and Woolford (1984) and Tweedie (1983).

From standard linear framework, we know that an integrated process \( X_t \) of order \( d \) \( (I(d)) \) is called cointegrated \( CI(d, b) \) if a linear combination denoted as \( \beta' X_t \), with \( \beta \neq 0 \), is \( I(d-b) \), where \( b = 1, \ldots, d \) and \( d = 1, \ldots \). However, as mention above, these definitions are not sufficient to handle all situations when dealing with nonlinear functions. As a result, Bec and Rahbek (2002) define nonlinear cointegration, in a Markov theory framework, as follows,

**Definition 6.** A process \( X_t \) in a nonlinear \( I(1) \)-type process if \( X_t - EX_t \) is not geometrically ergodic, while \( \Delta(X_t - EX_t) \) is.

Therefore, under this approach the concept of geometric ergodicity corresponds to the linear concept of integratedness. Using the definition above, the following Proposition is a result of Proposition 2,
**Proposition 3.** For a process \( X_t \) given in (6) with one of the transition functions specified in (7). If assumptions in Condition 2 and Proposition 2 hold, \( X_t \) is a nonstationary type process with \( \Delta X_t \) and \( \beta' X_t \) geometrically ergodic processes. Moreover, for initial distributions for \( \Delta X_t \) and \( \beta' X_t \), the processes are stationary and the cointegrating relationships are given by \( \beta' X_t \). Finally, \( \Delta X_t \) and \( \beta' X_t \) have finite moments up to order \( 2n \).

The proof of the Proposition above is presented in Appendix 1. Based on this result (assuming that \( \beta \) is known), standard CLT and LLN can be employed to derive asymptotic normality, consistency of M-estimators and inference with our nonlinear error correction model. Furthermore, following Corradi et al. (2000), the KPSS stationary test and Shin’s cointegration tests (Shin (1994)) are applicable (since the distribution under the null of stationarity can be derived from our results). Simultaneously, Proposition 3 can be seen as an illustration of linear stochastic comovement introduced by Corradi et al. (2000); in which they propose a consistent test which has a null of linear cointegration (comovement), and an alternative of “nonlinear” cointegration.

As observed in Pötscher and Prucha (1997) and Escribano and Mira (2000), stationary ergodicity can be seen as a strong assumption, since many economic data exhibit, both temporal dependence, a temporal heterogeneity. Therefore, stochastic and asymptotic properties of our standard nonlinear error correction model, under such conditions is analyzed in the next section using concepts of mixing and NED conditions. This approach represents the second strands in the literature in dynamic nonlinear asymptotic theory.

### 3.2 Representation Theory based on NED

The aim of the first part of this section is to introduce more general concepts than integrated series, that permit us to study a variety of stationarity processes in a nonlinear context. In linear time series, it is common to characterize a nonstationary process as a summation of a sequence of stationary process.

An integrated process is a particular type of nonstationary process, whose stochastic part is formed through the accumulation over time of stationary disturbances. Furthermore, under this type of (weak) nonstationarity, a integrated process has time-varying moments. Integrated processes and weak stationarity are suitable assumptions for analyzing linear time series models. However, Granger (1995) shows that these conditions are not sufficient to handle processes with nonlinear functions. As in the current literature on nonlinear models, this section relies on the concepts of ‘mixing conditions’ and ‘near-epoch’ dependence.

The concept of NED processes was introduced in the econometric literature by Gallant and White (1988) and Pötscher and Prucha (1997) for analyzing...
estimation and inference in dynamic nonlinear time series processes. Recently, Escribano and Mira (2000) extend the framework of Gallant and White (1988) by introducing explicit assumptions, such as moment conditions and properties on the nonlinear function, that guarantees the consistency and asymptotic normality of NLS. On the same line, Davidson (2001) obtains the (stability) conditions under which nonlinear models, such that GARCH, bilinear and threshold models, are NED process.

3.2.1 Some measures of persistence

The concept of mixing is appropriate to study the degree of dependence in the memory of a time series. The notion implies that as the time span between two events increases, the dependence between past and future events becomes negligible. Let \((X_h, h > 1)\) be a random sequence on \((\mathbb{R}^p, B^p, P)\), where \(\mathbb{R}^p\) defines the sample space, \(B^p\) is the Borel \(\sigma\)-algebra on \(\mathbb{R}^p\) and \(P\) a probability measure with domain \(B^p\). Consider a generated \(\sigma\)-algebra \(F^h_m = \sigma(X_t : m \leq t \leq h)\), then,

**Definition 3.** \(\{X_t\}\) is strong mixing (or \(\alpha\)-mixing), if

\[
\alpha(h) = \sup_{m \geq 1} \alpha(F^m_1, F^\infty_{m+h}) \xrightarrow{h \to \infty} 0
\]

where \(\alpha(h)\) is a statistical measures of the degree of dependence between events, such that for two events \(A \in F^m_1\) and \(B \in F^\infty_{m+h}\),

\[
\alpha(F^m_1, F^\infty_{m+h}) = \sup \{P(A \cap B) - P(A)P(B)\}
\]

Uniform and strict forms of mixing can be defined, however, for our purposes we shall only introduce the strong mixing case. \(\alpha(h)\) measures the speed of mixing or amount of dependence between events separated by \(h\) periods. If \(\alpha(h) = O(h^\lambda)\) for all \(\lambda < -\alpha\), then \(\alpha(h)\) is said to be of size \(-\alpha\).

Davidson (1994) and Escribano and Mira (2002) argue that the mixing concept has a serious drawback for time series analysis, in the sense that a function of a mixing sequence depending on a infinite number of lags may not be mixing. Only measurable functions of finitely elements of a mixing process are themselves mixing. Therefore, for static models, mixing conditions are adequate assumptions for modelling nonlinear specifications. However, as Pötscher and Prucha (1997) point out, in dynamic models the endogenous variables will typically be a function on the infinite history of exogenous variables and disturbances.

Furthermore, even a simple AR(1) process with i.i.d.disturbances can fail to be either \(\phi\)-mixing or \(\alpha\)-mixing (See theorem 14.7 in Davidson (1994)) . As a result the literature has replaced this condition by a weaker assumption, more easy to verify in empirical applications. This notion for nonlinear processes is
based on NED on mixing process. In which, it is possible to obtain useful results by considering functions of a possibly infinite history of an underlying mixing process, provided that the extent to which the function considered depends on a distant past or future of the underlying process. Denote $\| \cdot \|_r$ the $L_r$ random norm $\| \cdot \|_r = (E(|\cdot|^r))^{1/r}$.

**Definition 4.** \{ $X_t$ \}$_{t=\infty}^{-\infty}$, a stochastic sequence of random variables on a probability space ($\mathbb{R}^p, B^p, P$), is said to be $L_r$-NED on the sequence \{ $W_t$ \} of size $-a$, if for $r > 0$,

$$\| X_t - E (X_t / F_{t-m}^t) \|_r \leq d_t v_m$$

where $v_m \to 0$ as $m$ becomes large and \{ $d_t$ \}$_{t=\infty}^{+\infty}$ is a sequence of positive constants. Here $F_{t-m}^t = \sigma (W_{t-m}, \ldots, W_t)$, such that \{ $F_{t-m}^t$ \}$_{t=\infty}^{+\infty}$ is an increasing sequence of $\sigma$-fields. Then, $X_t$ is said to be $L_r$-NED on \{ $W_t$ \} of size $-a$ if $v_m = O(m^{-\lambda})$ for $\lambda < -a$.

Intuitively, this means that $X_t$ depends on recent values (epoch) of $W_t$, when $v_m$ becomes small (i.e. $X_t$ does not depend “too much” on the distant past or future). Formally, (15) can be seen as a nonparametric restriction on the memory of the process that nonetheless constrains only a sequence of low-order moments. Furthermore, by definition $E (X_t / F_{t-m}) = X_t$, hence if the whole history of \{ $W_t$ \} is known, $X_t$ can be predicted exactly. If $X_t$ is dependent on the 'near epoch' of $W_t$, the random variable measured by the norm $\| \cdot \|_r$ should be close to zero when $m$ gets large. With the mixing notion, the size of the sequence \{ $v_m$ \} is defined $-a$, with $v_m = O(m^{-\lambda})$ for $\lambda < -a$. This general definition was proposed by Andrews (1988), and it is extendedly used in the econometrics literature (e.g. Andrews (1993), Davidson (1994 and 2001) Drufénot and Mignon (2002), Escribano and Mira (2002), Gallant and White (1988) and Pötscher and Prucha (1997)). The definition used by Gallant and White (1988) also includes $E \{ X_t^2 \} < \infty$.

The NED concept introduced above is more stringent to the concept of $L_r$-approximability used in the probability literature (see details in Pötscher and Prucha (1997)). One main idea behind these concepts relies on the asymptotic vanishing property of the approximation error in different $L_r$ norms or in terms of probability measures.

Escribano and Mira (2002) observe that, under some regular conditions, the NED property is not invariant to arbitrary transformations of the process (i.e. functions of NED are NED). Moreover, $E (X_t / F_{t-m}^t)$ is a mixing process, since it is a finite-lag $F_{t-m}^t/B^p$-measurable function of a mixing process, hence NED implies that \{ $X_t$ \} is “approximately” mixing
However, as pointed out by Davidson (1994), NED should not be considered as an alternative to mixing conditions and without the assumption of an underlying mixing process; the NED is only a property of the mapping from \( \{W_t\} \) to \( \{X_t\} \), not of the stochastic sequence. By making the process NED a function of a mixing process, serial dependence is restricted sufficiently and subject to suitable restrictions on moment conditions, the process can be *mixingale*, an asymptotic analogue of a martingale (see Lemma 3.14 in Gallant and White (1988)). This result makes it straightforward to establish CLT, FCLT and LLN. Intuitively, LLN and CLT can be applied as long as the process \( f_{Wt} \) is sufficiently mixing and the nonlinear functions has a “fading memory” such that they show “declining weights” on high lags of \( \{W_t\} \). With NED processes, the notion of “declining weights” is formalized by approximating functions that depends on infinitely lags of \( \{W_t\} \) by functions that only depends on finitely elements of \( \{W_t\} \).


### 3.2.2 Representation Theory under NED conditions

A seminal work in generalizing Granger representation theorem in a nonlinear framework using NED conditions is developed in Escribano and Mira (2002). Similar to our results in Theorem 3, they extend partially the theorem using equation (4) with \( k = 2 \). Roughly speaking, the assumptions given in Condition 2 are replaced by the following three main ideas. Firstly, the map \( F (\beta'X_t) \) where \( F : \mathbb{R}^r \rightarrow \mathbb{R}^p \) is assumed to be continuously differentiable and satisfies general Lipschitz conditions. Second, the largest eigenvalue (spectral radius) of the matrix of first partial derivatives is smaller that 1. This boundedness condition plays a similar role to the drift function in the Markov chain theory, and it can be seen as a *contraction mapping* condition. thirdly, the sequences \( \varepsilon_t \) is a \( \alpha \)-mixing process with \( E \| \varepsilon_t \|_S^2 < \infty \).

However, as above our proof concentrates on equation (9) and our aim is to obtain a nonlinear stationary-type process using NED conditions, similar to Proposition 3. In contrast to the previous section, the result follows with more primitive disturbances, though the remaining assumptions do not vary, in particular our stability condition, given in Assumption (2b). Therefore, similar to the Assumption MX in Gallant and White (1988) and Escribano and

---

3Note that Escribano and Mira (2002) use more general \( L_p \) random norm such as \( \|X\|_S = (E (\|X\|_S^p)^{1/p} \) in which \( \|X\|_S \) is a Minkowski \( S \)-norm (i.e. for a \( (m \times n) \) matrices \( A \), \( \|A\|_S = [\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^S]^{1/S} \) for \( 1 \leq S < \infty \)).
Mira (2000), combined with Assumption LR Escribano and Mira (2000), our Assumption (1b) can be relaxed as follows,

**Assumption 1b'.** \( \{ \eta_t^* \} \) is a \( \alpha \)-mixing sequence of size \(-r/(2(r-1))\) for \( r \geq 2 \) and \( E \| \eta_t^* \|^r \leq \Delta_{r}^{(r)} < \infty \).

We should point out that the close links between mixing (and/or NED) and geometric ergodicity are relatively well known in the statistics literature. As in Carrasco (2002) (see proof of Proposition 3), the results of Doukhan (1994) can be extended in our Proposition 3 implying that \( X_t \) and \( \eta_t^* \) follow NED processes; roughly speaking, ergodicity may imply mixing conditions in a stochastic process. However, our aim here show the condition under these two seemingly separate strands of asymptotic theory present in econometrics literature can be linked. That is, replace our Assumption (1b) by a less stringent condition specified above and maintain the stability condition that ensures both geometric ergodicity and NED.

As before, our extension of Granger’s representation theorem in a NED context requires a separation of “nonlinear” stationary and nonstationary processes, thus,

**Definition 6'.** A process \( X_t \) in a nonlinear \( I(1) \)-type process if \( X_t - EX_t \) is not NED on a \( \alpha \)-mixing process, while \( \Delta (X_t - EX_t) \) is.

A similar concept is given in Escribano and Mira (2000). Thus, we are able to write down a partial extension of Granger’s representation theorem in a NED environment, as follows,

**Proposition 3'.** For a process \( X_t \) given in (6) with one of the transition functions specified in (7). If assumptions in (1a), (1b'), (2a) and (2b) hold, \( X_t \) is a nonstationary type process with \( \Delta X_t \) and \( \beta' X_t \) being a \( L_2 - \text{NED} \) on a \( \alpha \)-mixing sequence \( \{ \eta_t^* \} \); and the cointegrating relationships are given by \( \beta' X_t \).

See Appendix 2 for the proof. Once more, the analysis concentrates in the companion form of equation (9), i.e.

\[
V_t = (\Phi_2 + s_{t-1} (\Phi_1 - \Phi_2)) V_{t-1} + \eta_t^* \tag{16}
\]

For notational issues redefine the time index on \( s_t \), such that \( s_{t-1} = s (\gamma V_{t-1}) \). By standard repeated substitution up to time \( m + 1 \) and with the convention that an empty product equals to one \( (\prod_{i=k}^l A_i = 1, \text{if} \ l < k) \), the equation above can be rewritten as follows,

\[
V_t = \sum_{j=0}^{m} \prod_{k=1}^{j} (\Phi_2 + s_{t-k} (\Phi_1 - \Phi_2)) \eta_{t-j}^* + \prod_{k=1}^{m+1} (\Phi_2 + s_{t-k} (\Phi_1 - \Phi_2)) V_{t-m-1} \tag{17}
\]

\(^4\)For simplicity, our measure of distance is based on euclidean distances as before, there \( S = 1 \) is assumed.
The $L_r - \text{NED}$ properties of the process in (17) can be analyzed in terms of the existence of a finite lag approximator to the process, such that we build a $F_{t-m}^{m}/B$-measurable function such that $\tilde{V}_{t}^{m} : \mathbb{R}^{2p(2m+1)} \to \mathbb{R}^{2p}$, in which $F_{t-m}^{m} = \sigma(\eta_{t-m}, \ldots, \eta_{t+m})$. Consequently, equation (17) can be approximated by a measurable function that considers the $m$ recent past disturbances (epochs), i.e. setting $\eta_{t-j} = 0$ for $j > m$,

$$\tilde{V}_{t}^{m} = \sum_{j=0}^{m} \prod_{k=1}^{j} (\Phi_{2} + \tilde{s}_{t-k} (\Phi_{1} - \Phi_{2})) \eta_{t-j}$$

where $\tilde{s}_{t-k} = s(\tilde{V}_{t-k}^{m-k})$ (i.e. $\tilde{s}_{t-k} = 1\left(\tilde{\gamma}^{m-k} > c^{*}\right)$ for the SETAR case $\tilde{s}_{t-k} = G\left(\gamma^{m-k}, \gamma^{*}, c^{*}\right)$) for the STAR-type process, in which $c^{*} = (c,0,0,0)^{t}$. The approach intends to show that difference between the process in (17) and its $F_{t-m}^{m}/B$-measurable approximation given above becomes asymptotically negligible measured by a suitable random norm $\|\cdot\|_{2r}$. Note that the conditional mean $E(X_{t} / \tilde{F}_{t-m}^{m})$ can be thought as a $F_{t-m}^{m}/B$-measurable function and if the underlying process $\{\eta_{t}\}$ is a mixing sequence then both the approximation and the conditional expectation preserve this property. Hence, using a $L_{2}$-random norm can be written as (details are given in Appendix 2),

$$\left\| V_{t} - \tilde{V}_{t}^{m} \right\|_{2} \leq C_1 \sum_{j=1}^{m} \sum_{k=j}^{m} \rho^{\frac{1}{2}(m+j-2)} \cdot M_{2} \left\| V_{t-j} - \tilde{V}_{t-j}^{m-j} \right\|_{2}^{\frac{r-1}{2}} + \rho^{\frac{r}{2}} \cdot M_{1} \left\| V_{t-m} \right\|_{2r}$$

(18)

where $C_{1} = 2^{\frac{r}{2}} B^{\frac{r-2}{2}} \Delta_{n}^{(r)}$.

$$\left\| \Lambda_{k=1}^{m+1} \right\|_{2} < \max_{i=1,2} \rho \left( \Phi_{i}^{t} \otimes \Phi_{i}^{t} \right)^{\frac{r}{2}} \cdot M_{1} \to 0 \text{ as } m \to \infty$$

(19)

and,

$$\left\| \tilde{\Lambda}_{i=1}^{j-1} (\Phi_{1} - \Phi_{2}) \Lambda_{i=1}^{j+1} \right\|_{2} < \max_{i=1,2} \rho \left( \Phi_{i}^{t} \otimes \Phi_{i}^{t} \right)^{\frac{1}{2}(m+j-2)} \cdot M_{2} \to 0 \text{ as } m \to \infty$$

(20)

The product of matrices are defined as, $\Lambda_{k=1}^{u} = \prod_{k=1}^{u} \left( \Phi_{2} + s_{t-k} (\Phi_{1} - \Phi_{2}) \right)$ and $\tilde{\Lambda}_{i=1}^{u} = \prod_{k=1}^{u} \left( \Phi_{2} + \tilde{s}_{t-k} (\Phi_{1} - \Phi_{2}) \right)$. Firstly, similar to Davidson (2002), the result given above follows from Lemma A.2.1. (Appendix 2) in which a Lipschitz condition is additionally assumed on the transition function, meaning that $\|s(v_{1}) - s(v_{2})\| \leq B \|v_{1} - v_{2}\|$ for all $v_{1}$ and $v_{2}$ and some $B > 0$. Secondly, the geometric convergence of the random processes in (19) and (20) are analyzed in Lemma A.2.2. (Appendix 2) using the arguments from Pham (1985).
and 1986) and Tjøstheim (1990) based on analyzing symmetric tensor product spaces and the stability condition given in Assumption (2b) (i.e. $\rho(\cdot) < 1$).

As a result, the first part of Proposition 3' (partial extension of Granger’s Representation theorem) holds, since the stochastic process in $(V_t, t \geq 1)$ is a $L_2 - NED$ of on the underlying $\alpha-$mixing sequence $\{\eta_t^*\}$ as,

$$\|V_t - E (V_t | F_{t-m})\|_2 \leq \left\|V_t - \hat{V}_t^m\right\|_2 \rightarrow 0 \text{ as } m \rightarrow \infty$$

or $\left\|V_t - \hat{V}_t^m\right\|_2$ has an order of $O \left(\max_{i=1,2} \rho \left(\Phi_i \otimes \Phi_i^*\right)^m\right)$. Furthermore, if the sequence $(V_t, t \geq 1)$ is assumed to be $r-$uniformly integrable, i.e. $\|V_t\|_r \leq \Delta_v < \infty$, then Lemma 3.14 in Gallant and White (1988) and Theorem 17.5 in Davidson (1994) holds, therefore $\{V_t, F_t \}$ is a $L_2 - mixingale$ of size $b$, where $b$ is the size of $L_2 - NED$ process. The last result is crucial for developing asymptotic theory for NLS or GMM estimators, as in Pötscher and Prucha (1997) and Gallant and White (1988), since a battery of (U)LLN and (F)CLT are available for mixingale process.

4 Conclusions

The paper partially extends Granger representation theorem for a specific class of nonlinear models, i.e. (S)TAR specifications. The idea of stationarity in this nonlinear cointegrated system relies on the concepts of geometric ergodicity and near epoch dependence, both notions of stochastic stability have been used to develop asymptotic theory of nonlinear dynamic models.

In the first part of the paper, Markov chain tools are used to show that Vector Error Correction model with nonlinear adjustments is govern by a geometric ergodic process. This result is based on strong assumptions regarding the distribution and moments of the disturbances. In the second part, we relax these assumptions and with more primitive disturbances we control the memory of $\Delta X_t$ and $\beta' X_t$ so that they follow a near epoch dependent processes. These results help link these two seemingly separate strands of asymptotic theory present in econometrics literature, which have been already closely linked in the statistics literature (i.e. an ergodicity implies mixing conditions in a stochastic process). Furthermore, the connection between these two approaches rests on stability conditions that ensure both ergodicity and NED.

Moreover, the understanding of stochastic stability tend to differ between the two approaches. In Markov chain theory, stability is conceived as a convergence to the “centre” of the space (i.e. the attractor set). By using NED, stability is closely linked with controlling the memory of the process.
5 Proof Proposition 3 (appendix 1)

Premultiplying (9) by \((\beta_\perp, \beta_\perp')\)', one can obtain,
\[
\beta'X_t = s_t \left[ (I_\perp + \beta'\alpha) \beta'X_{t-1} + \beta'\Gamma_1 \Delta X_{t-1} \right] \\
+ (1 - s_t) \left[ (I_\perp + \beta'\alpha) \beta'X_{t-1} + \beta'G_1 \Delta X_{t-1} \right] + \beta'\varepsilon_t
\]
and,
\[
\beta_\perp' \Delta X_t = s_t \left[ \beta_\perp' \alpha \beta'X_{t-1} + \beta_\perp' \Gamma_1 \Delta X_{t-1} \right] \\
+ (1 - s_t) \left[ \beta_\perp' \alpha \beta'X_{t-1} + \beta_\perp' G_1 \Delta X_{t-1} \right] + \beta_\perp' \varepsilon_t
\]
Since \(\Delta X_t = \beta_\perp \Delta Z_t + \beta_\parallel Z_t^{\perp 1}, \) where \(\beta = (\beta'\beta)^{-1}\) and \(\beta_\perp + \beta_\parallel \beta_\perp' = I_p,\)
the equation above can be collected in a matrix form as,
\[
V_t = s_t(\Phi_{11}V_{t-1} + \Phi_{12}V_{t-2}) + (1 - s_t) (\Phi_{21}V_{t-1} + \Phi_{22}V_{t-2}) + \eta_t \quad (A.1.)
\]
with \(p \times p\) elements \(\Phi_{ij}, \ i, j = 1, 2,\) given by,
\[
\Phi_{11} = \begin{pmatrix}
I_\perp + \beta'\alpha + \beta'\Gamma_1 \beta_\parallel \\
\beta_\perp' \alpha + \beta_\perp' \Gamma_1 \beta_\parallel \\
\beta_\perp' \alpha + \beta_\perp' \Gamma_1 \beta_\parallel
\end{pmatrix},
\Phi_{12} = \begin{pmatrix}
-\beta'\Gamma_1 \beta_\parallel & 0 \\
-\beta_\perp' \Gamma_1 \beta_\parallel & 0
\end{pmatrix},
\Phi_{21} = \begin{pmatrix}
I_\perp + \beta'\alpha + \beta'G_1 \beta_\parallel \\
\beta_\perp' \alpha + \beta_\perp' G_1 \beta_\parallel \\
\beta_\perp' \alpha + \beta_\perp' G_1 \beta_\parallel
\end{pmatrix},
\Phi_{22} = \begin{pmatrix}
-\beta'G_1 \beta_\parallel & 0 \\
-\beta_\perp' G_1 \beta_\parallel & 0
\end{pmatrix}
\]
The disturbances \(\eta_t = (\beta_\perp, \beta_\parallel')'\varepsilon_t,\) are governed by \(\eta_t \sim i.i.d. (0, \Sigma), \Sigma = (\beta, \beta_\perp)' \Omega (\beta, \beta_\perp) > 0\) as a consequence of Assumption 1b. Equation (A.1.)
have to be rewritten in a companion form, in order to obtain a homogeneous
Markov chain on \(V_t = (V_t', V_t^{\perp 1})'.\) Thus, (A.1.) can be written as follows,
\[
V_t = s_t \Phi_i V_{t-1} + (1 - s_t) \Phi_2 V_{t-1} + \eta_t^* \quad (A.2.)
\]
where, \(A = \begin{pmatrix}
\Phi_{11} & \Phi_{12} \\
I_p & 0
\end{pmatrix}, \ B = \begin{pmatrix}
\Phi_{21} & \Phi_{22} \\
I_p & 0
\end{pmatrix}, \ \eta_t^* = (I_p, 0)' \eta_t\) and
covariance matrix given by \(\Sigma^* = \beta'\Omega \beta.\) Here, the coefficients matrices \(\Phi_1\) and \(\Phi_2\) have dimension of \(p \times 2p.\) Simultaneously, the support of the transition
function \(s : \mathbb{R}^p \rightarrow [0, 1]\) has to be changed from \(\beta'X\) to \(V_t,\) therefore a matrix
\(\gamma = (I_p, 0, 0)'\) can be defined so \(s_t = p (\gamma'V_{t-1})'.\)

As mention above, Assumption (1b) on \(\{\eta_t^*\}\) implies that the chain \((V_t, t \geq 1)\)
is \(\phi\text{-irreducible and aperiodic}.\) Furthermore, as the nonlinear structure of (11)
is given by piecewise linear functions \(f(\cdot)\) is a compact function and since by
Assumption (1b) \(\eta_t^*\) has a continuous density, then the existence of small
compact sets \(C\) on \(\mathbb{R}^{2p}\) is guarantee. Details proofs can be found in Chan and Tong
(1985) and (1986). Our proof uses the following drift criteria,
\[
g(v) = 1 + 1_{[\gamma'v > 0]} v'D_1v + 1_{[\gamma'v \leq 0]} v'D_2v \geq 1 \quad (A.3.)
\]
\[
D_j = \sum_{i=0}^{\infty} \Phi_j^i \Phi_j^i = (I_{2p} - \Phi_j' \Phi_j)^{-1}, \text{ for } j = 1, 2 \quad (21)
\]

21
The first inequality holds for all \( v \in \mathbb{R}^{2p} \), since \( D_j \) is a positive definite matrix, for \( j = 1, 2 \). Furthermore, the ultimate identity holds by assumption 2b (i.e. \( \sup_{j=1,2} \rho(\Phi_j) < 1 \implies \rho(\Phi_j, \Phi_j) < 1 \), for \( j = 1, 2 \)), as a result,

\[
g(V_t) = 1 + \sum_{j=1}^{2} V_j^t D_j V_t
\]

\[
= 1 + \sum_{j=1}^{2} (s_t \Phi_1 V_{t-1} + (1 - s_t) \Phi_2 V_{t-1} + \eta_j^t)^t D_j
\]

where, \( j = 1 \) when \( \gamma^t v > 0 \), and \( j = 2 \) in the opposite case. Taking conditional expectation with respect to \( V_{t-1} \) to build the drift criteria, and given that \( \eta_j^t \) is independent of \( V_{t-1} \) by assumption (1), one finds,

\[
E [g(V_t) \mid V_{t-1} = v] = 1 + \sum_{j=1}^{2} (s_v^2 \Phi_1^t D_j \Phi_1 v + 2s_v (1 - s_v) \Phi_1^t D_j \Phi_2 v + (1 - s_v)^2 \Phi_2^t D_j \Phi_2 v + v + tr (D_j \beta^t \Omega \beta))
\]

Firstly, let assume \( \gamma^t v > 0 \) so a bounded attraction set can be defined as \( C = \{ v \in \mathbb{R}^{2p} : \gamma^t D_1 v \leq M_1 \} \), where \( M_1 > 1 \) is a appropriate constant. Let suppose that the chain remains outside the attraction centre \( C \) (i.e. \( v \in C^c \)), so the process tends to explode positively (since \( \gamma^t v > 0 \), such that \( v \to \infty \) or \( v^t D_1 v \to \infty \). By assumption (1a), the transition function will behave like \( s_v \to 1 \) as \( v \to \infty \). Recalling the definition of \( D_1 \), it can be shown that \( \Phi_1^t D_1 \Phi_1 = (\Phi_1 - I_{2p}) \) and \( \Phi_1^t D_1 \Phi_1 v = (\gamma^t \times \gamma^t) (\Phi_1^t \times \Phi_1^t) vec(D_1) \), therefore, the equation above can be written as,

\[
E [g(V_t) \mid V_{t-1} = v] = 1 + \gamma^t D_1 v - \gamma^t v + tr (D_1 \beta^t \Omega \beta)
\]

\[
= 1 + (\gamma^t \times \gamma^t) vec(D_1) - (\gamma^t \times \gamma^t) + tr (D_1 \beta^t \Omega \beta)
\]

to conclude that,

\[
E [g(V_t) \mid V_{t-1} = v] = g(v) - (\gamma^t \times \gamma^t) + tr (D_1 \beta^t \Omega \beta)
\]

\[
= g(v) \left( 1 - \frac{(\gamma^t \times \gamma^t) - tr (D_1 \beta^t \Omega \beta)}{g(v)} \right)
\]

Since \( \gamma^t D_1 v \setminus M_1 > 1 \),

\[
g(v) = 1 + \gamma^t D_1 v = 1 + (\gamma^t \times \gamma^t) vec(D_1)
\]

\[
\leq \frac{(\gamma^t \times \gamma^t) vec(D_1)}{M} + (\gamma^t \times \gamma^t) vec(D_1)
\]

\[
\leq \frac{1 + M_1}{M_1} (\gamma^t \times \gamma^t) vec(D_1) \leq 2 (\gamma^t \times \gamma^t) vec(D_1)
\]

22
So that in this case,

$$E \left[ g(V_t) \mid V_{t-1} = v \right] \leq g(v) \left( 1 - \frac{(v' \otimes v') - \text{tr} \left( D_1 \beta' \Omega \beta \right)}{2(v' \otimes v') \text{vec}(D_1)} \right)$$

$$\leq g(v) \left( 1 - \frac{1}{2 \text{vec}(D_1)} + \frac{\text{tr} \left( D_1 \beta' \Omega \beta \right)}{2(v' \otimes v') \text{vec}(D_1)} \right)$$

Based on our original model, the numerator of the last component inside the brackets is a finite number, thus its asymptotic behavior implies that $\text{tr} \left( D_1 \beta' \Omega \beta \right) = o(n)$ as $\|v\| \to \infty$, and therefore, outside the attraction set,

$$E \left[ g(V_t) \mid V_{t-1} = v \right] \leq g(v) \left( 1 - \frac{1}{2 \rho(D_1)} \right) \leq g(v) \beta(v)$$

By assumption (2b), when $v' D_1 v > M_1$, that is $v \in C^c$, the conditional expected value of $g(V_t)$ at most smaller than $g(v)$ at $V_{t-1} = v$ since $0 < \beta(v) < 1$. As a consequence, as if the process $\{v_t\}_{t=1,2,\ldots}$ tends to explode overtime located on $C^c$, the chain will dissipate energy through the generalized energy function $g(\cdot)$ until process is forced back to the attraction set $C$.

On $C$, $v' D_1 v$ is bounded by $M_1$ for all $v \in \mathbb{R}^p$, hence $E \left[ g(V_t) \mid V_{t-1} = v \right]$ is continuous and bounded on a compact set $C$. Similar analysis holds for the case when $v' \leq 0$ and $v \to -\infty$, so at $v \in C^c$ the process is diverging such as $v' D_2 v > M_1$. In this situation, $\rho(\Phi_2) < 1$ (based on Assumption 2b) will ensure the convergence towards the attraction set $C$. Subsequently, the proof is complete for the $\|v\|$-geometric ergodicity on the Markov chain $\{V_t\}_{t=1,2,\ldots}$. Moreover, following Theorem 1, if the chain is started with initial distribution $\pi$, i.e. $\ell(v_0) = \pi$, then $(V_t, t \geq 1)$ is strict stationary.

Similar to Theorem 2.2. in Petruccelli and Woolford (1984), this result implies that the drift function $E_\pi g(V_t) < \infty$ is integrable with respect to $\pi$, therefore the chain $(V_t, t \geq 1)$ has finite second moments, given the quadratic structure of our drift criteria and Assumption (2a) (i.e. $E |\eta_i|^2 < \infty$ for $n = 1$). As pointed out by Bec and Rahbek (2001), higher moments can be obtained by modifying the structure of the drift criteria function and set $n \geq 1$. The proof is developed in Lemma 6.1 in Tjøstheim (1990), Theorem 2.2. in Petruccelli and Woolford (1984) and Tweedie (1983). □

6 Proof Proposition 3'(appendix 2)

As discussed above, we need to prove that $(V_t, t \geq 1)$ is $L_2$-approximable on the mixing sequence $\{\eta_1^2\}$; that is, the difference between $V_t$ and $\tilde{V}_t^m$ is asymptotically negligible measured by a suitable random norm, therefore substracting
equation (24) from (17),
\[ V_t - \hat{V}_t^m = \sum_{j=1}^{m} \left( \prod_{k=1}^{j} (\Phi_2 + s_{t-k} (\Phi_1 - \Phi_2)) - \prod_{k=1}^{j} (\Phi_2 + \hat{s}_{t-k} (\Phi_1 - \Phi_2)) \right) \eta_{t-j}^* \]
\[ + \prod_{k=1}^{m+1} (\Phi_2 + s_{t-k} (\Phi_1 - \Phi_2)) V_{t-m-1} \]

After some algebra and rearranging certain arrays, as in Davidson (2002), the equation above can be rewritten as,

\[ \|V_t - \hat{V}_t^m\|_2 \leq \sum_{j=1}^{m} \left( \sum_{k=j}^{m} \left( \hat{A}_{i=1}^{k} (\Phi_1 - \Phi_2) A_{i=j+1}^{k} \eta_{t-k}^* \right) \left( (s_{t-j} - \hat{s}_{t-j}) \right) \right) \]
\[ + \left\| \sum_{k=1}^{m+1} V_{t-m-1} \right\|_2 \]

where \( A_{i=1}^{k} = \prod_{k=1}^{w} (\Phi_2 + s_{t-k} (\Phi_1 - \Phi_2)) \) and \( \hat{A}_{i=1}^{w} = \prod_{k=1}^{w} (\Phi_2 + \hat{s}_{t-k} (\Phi_1 - \Phi_2)) \).

Next, a \( L_2 \)-random norm is applied on the equation above and by Minkowski’s inequality (i.e. \( \sum_{i=1}^{m} X_i \|_r \leq \sum_{i=1}^{m} \|X_i\|_r \)),

\[ \|V_t - \hat{V}_t^m\|_2 \leq \sum_{j=1}^{m} \left( \sum_{k=j}^{m} \left( \hat{A}_{i=1}^{k} (\Phi_1 - \Phi_2) A_{i=j+1}^{k} \eta_{t-k}^* \right) \left( (s_{t-j} - \hat{s}_{t-j}) \right) \right) \]
\[ + \left\| \sum_{k=1}^{m+1} V_{t-m-1} \right\|_2 \]

Applying Minkowski’s inequality and Cauchy-Shwartz’s inequality (i.e. \( E (XY)^2 \leq E (X)^2 E (Y)^2 \)) to the first component in the right hand side of the inequality above. And using Hölder’s inequality (i.e. for any \( p, E |XY| \leq E^{1/p} |X|^p E^{1/q} |X|^q \) for \( p = r \) and \( q = r/(r-1) \)) on the second component, we obtain,

\[ \|V_t - \hat{V}_t^m\|_2 \leq \sum_{j=1}^{m} \sum_{k=j}^{m} \left( \hat{A}_{i=1}^{k} (\Phi_1 - \Phi_2) A_{i=j+1}^{k} \eta_{t-k}^* \right) \left( (s_{t-j} - \hat{s}_{t-j}) \right) \]
\[ + \left\| \sum_{k=1}^{m+1} V_{t-m-1} \right\|_2 \]
\[ \leq \sum_{j=1}^{m} \sum_{k=j}^{m} \left( \hat{A}_{i=1}^{k} (\Phi_1 - \Phi_2) A_{i=j+1}^{k} \eta_{t-k}^* \right) \left( (s_{t-j} - \hat{s}_{t-j}) \right) \]
\[ + \left\| \sum_{k=1}^{m+1} V_{t-m-1} \right\|_2 \]

(24)

Now, we need to prove the following claim,

**Lemma A2.1.** Assuming that Lipschitz condition holds for the transition function \( s_{t-k} = s (\sqrt{V_{t-k}}) \), so \( \|s (v_1) - s (v_2)\| \leq B \|v_1 - v_2\| \) for all \( v_1 \) and
\( v_2 \) and some \( B > 0 \). Then, the inequality in (??) holds, i.e. \( \| s_{t-j} - \hat{s}_{t-j} \|_2 \leq 2 \rho_1^{\frac{1}{r+1}} B \frac{r}{r+1} \| V_{t-j} - \hat{V}_{t-j} \|_2^{\frac{r}{r+1}} \).

Proof of Lemma A2.1. The arguments that we need follow from Theorem 17.13 in Davidson (1994). First, the Lipschitz condition can be reexpressed as,

\[
\| s_{t-j} - \hat{s}_{t-j} \| \leq \min \left\{ B \left\| V_{t-j} - \hat{V}_{t-j} \right\|, 2 \right\}
\]

(25)

Let's define \( Z = B \left\| V_{t-j} - \hat{V}_{t-j} \right\| / 2 \), so it follows that \( \| s_{t-j} - \hat{s}_{t-j} \| \leq 2 \min (Z, 1) \) and for any \( r > 1 \),

\[
E \| s_{t-j} - \hat{s}_{t-j} \|^{\frac{2(r+1)}{r+1}} = \int_{Z \leq 1} \| s_{t-j} - \hat{s}_{t-j} \|^{\frac{2(r+1)}{r+1}} dP + \int_{Z > 1} \| s_{t-j} - \hat{s}_{t-j} \|^{\frac{2(r+1)}{r+1}} dP
\]

\[
\leq 2 \int_{Z \leq 1} Z^{\frac{2(r+1)}{r+1}} dP + 2 \int_{Z > 1} Z^{\frac{2(r+1)}{r+1}} dP
\]

where \( P \) is a probability measure. Since \( r > 1 \), the ratio \( (2(r+1)) / (r-1) \geq 2 \). Therefore, if \( Z \leq 1 \), \( |Z|^{\frac{2(r+1)}{r+1}} \leq |Z|^2 \), then,

\[
E \| s_{t-j} - \hat{s}_{t-j} \|^{\frac{2(r+1)}{r+1}} \leq 2 \int_{Z \leq 1} Z^2 dP + \int_{Z > 1} Z^2 dP
\]

\[
= 2 \int_{Z \leq 1} Z^2 dP + 2 \int_{Z > 1} Z^2 dP
\]

Hence, using the definition of \( Z \) leads to,

\[
E \| s_{t-j} - \hat{s}_{t-j} \|^{\frac{2(r+1)}{r+1}} \leq 2 \rho_1^{\frac{1}{r+1}} B^2 \left\| V_{t-j} - \hat{V}_{t-j} \right\|^2
\]

(26)

Note that by Lyapunov’s inequality (i.e. \( \| \cdot \|_2 \leq \| \cdot \|^{\frac{r+1}{r}} \)), our results follows,

\[
\| s_{t-j} - \hat{s}_{t-j} \|_2 \leq \left( E \| s_{t-j} - \hat{s}_{t-j} \|^{\frac{2(r+1)}{r+1}} \right)^{\frac{r-1}{r+1}} \leq 2 \rho_1^{\frac{1}{r+1}} B \frac{r}{r+1} \| V_{t-j} - \hat{V}_{t-j} \|_2^{\frac{r}{r+1}}
\]

\[
\square
\]

Using the result in (??) proved in Lemma A2.1 and Hölder’s inequality, equation (??) (Eq. (??) in Section XX) can be rewritten as follows,
\[ \|V_t - \hat{V}_t\|^m \leq 2^{\tau/\tau} B^{\tau/\tau} \sum_{j=1}^m \sum_{k=j}^m \|\hat{\Lambda}_{i=1}^{j-1} (\Phi_1 - \Phi_2) \Lambda_{i=j+1}^k \|_{2r/(r-1)} \|\eta_{i-j}^*\|_{2r} \\
arrow \|V_{t-j} - \hat{V}_{t-j}\|_{2r/(r-1)} \|\Lambda_{k=1}^{m+1}\|_{2r/(r-1)} \|V_{t-m-1}\|_{2r} \\
arrow C_1 \sum_{j=1}^m \sum_{k=j}^m \|\hat{\Lambda}_{i=1}^{j-1} (\Phi_1 - \Phi_2) \Lambda_{i=j+1}^k \|_{2r/(r-1)} \|\Lambda_{k=1}^{m+1}\|_{2r/(r-1)} \|V_{t-m-1}\|_{2r} \\
arrow + \|\Lambda_{k=1}^{m+1}\|_{2r/(r-1)} \|V_{t-m-1}\|_{2r} \\
\]

where, we define \( C_1 = 2^{\tau/\tau} B^{\tau/\tau} \Delta_0^{(r)} \) using the moment condition of Assumption (1b'). The following lemma will be useful to understand the behavior of \( \|\Lambda_{k=1}^{m+1}\|_{2r/(r-1)} \) and \( \|\hat{\Lambda}_{i=1}^{j-1} (\Phi_1 - \Phi_2) \Lambda_{i=j+1}^k \|_{2r/(r-1)} \), and it is based on analyzing symmetric tensor product spaces as in Theorem 4.2 in Pham (1985), Lemma 2 in Pham (1986) and Section 5 in Tjøstheim (1990).

**Lemma A2.2.** Given assumptions (1a), (2a) and in particular the stability condition specified in (2b),

\[ \|\Lambda_{k=1}^{m+1}\|_{2r/(r-1)} < \max_{i=1,2} \rho (\Phi_i \otimes \Phi_i)^{m/2} M_1 \to 0 \text{ as } m \to \infty \]

and,

\[ \|\hat{\Lambda}_{i=1}^{j-1} (\Phi_1 - \Phi_2) \Lambda_{i=j+1}^k \|_{2r/(r-1)} < \max_{i=1,2} \rho (\Phi_i' \otimes \Phi_i')^{1/2(m+j-2)} M_2 \to 0 \text{ as } m \to \infty \]

where \( k, j = 1, 2, \ldots, m \), and \( M_1 \) and \( M_2 \) are nonnegative constants. As before, we define \( \Lambda_{k=1}^m = \prod_{k=1}^m (\Phi_2 + \hat{s}_{t-k} (\Phi_1 - \Phi_2)) \) and \( \hat{\Lambda}_{k=1} = \prod_{k=1}^m (\Phi_2 + \hat{s}_{t-k} (\Phi_1 - \Phi_2)). \)

**Proof of Lemma A2.2.** First, we concentrate on the sequence given by 
\[ \|\Lambda_{k=1}^{m+1}\|_{2r/(r-1)} = \|\prod_{k=1}^{m+1} (\Phi_2 + \hat{s}_{t-k} (\Phi_1 - \Phi_2))\|_{2r/(r-1)}. \]

Observe that,

\[ \Lambda_{k=1}^{m+1} = \prod_{k=1}^{m+1} (\Phi_2 + \hat{s}_{t-k} (\Phi_1 - \Phi_2)) = \Phi_{t-1}^* \Lambda_{k=2}^{m+1} \]

where \( \Phi_{t-1} = (\Phi_2 + \hat{s}_{t-1} (\Phi_1 - \Phi_2)). \) By the definition of a Euclidean norm, we can obtain for this sequence,

\[ E \left( \|\Lambda_{k=1}^{m+1}\|^2 / F_{t-m-1} \right) = E \left( \text{Tr} \left( (\Lambda_{k=1}^{m+1})' \Lambda_{k=1}^{m+1} / F_{t-m-1} \right) \right) \]
\[ = \text{Tr} \left( E \left( (\Lambda_{k=1}^{m+1})' / F_{t-m-1} \right) \right) \]
\[ = \text{Tr} \left( E \left( \Phi_{t-1}^* \Lambda_{k=2}^{m+1} (\Phi_{k=2}^{m+1})' (\Phi_{t-1}^*)' / F_{t-m-1} \right) \right) \]
The second equality follows from \( E[\text{Tr}(A)] = \text{Tr}[E(A)] \) and \( \text{Tr}(AB) = \text{Tr}(BA) \), and by using equation (27) the last equation is obtained. Thus, if \( W_{k=1}^{m+1} = \Lambda_{k=1}^{m+1} (\Lambda_{k=1}^{m+1})' \) is defined, we have the result above can be written as,

\[
E \left( \| \Lambda_{k=1}^{m+1} \|^2 / F_{t}^{-m-1} \right) = \text{Tr} \left( E \left( \Phi_{t-1}^{*} W_{k=2}^{m+1} (\Phi_{t-1}^{*})' / F_{t}^{-m-1} \right) \right) \tag{28}
\]

Next, note that if the filtration \( F_{t}^{-m-1} \subseteq F_{t}^{-2} \), a Tower property of conditional expectations can be applied above such that,

\[
E \left( \Phi_{t-1}^{*} W_{k=2}^{m+1} (\Phi_{t-1}^{*})' / F_{t}^{-m-1} \right) = E \left[ E \left( \Phi_{t-1}^{*} W_{k=2}^{m+1} (\Phi_{t-1}^{*})' / F_{t}^{-2} \right) / F_{t}^{-m-1} \right] \tag{29}
\]

Furthermore, \( E \left( \Phi_{t-1}^{*} W_{k=2}^{m+1} (\Phi_{t-1}^{*})' / F_{t}^{-2} \right) \) and equals to,

\[
= E \left( (\Phi_{1} s_{t-1} + (1 - s_{t-1}) \Phi_{2}) W_{k=2}^{m+1} (\Phi_{1} s_{t-1} + (1 - s_{t-1}) \Phi_{2})' / F_{t}^{-2} \right)
= \Phi_{1} W_{k=2}^{m+1} \Phi_{1}' E \left( s_{t-1}^2 / F_{t}^{-2} \right) + \Phi_{2} W_{k=2}^{m+1} \Phi_{2}' E \left( (1 - s_{t-1})^2 / F_{t}^{-2} \right)
+ (\Phi_{1} W_{k=2}^{m+1} \Phi_{2}' + \Phi_{2} W_{k=2}^{m+1} \Phi_{1}') E \left( s_{t-1} (1 - s_{t-1}) / F_{t}^{-2} \right)
\]

Based on Assumption (1a), we know that the transition function for the (SE)TAR process is specified as an indicator function such that, \( s_{t-1} : \mathbb{R}^{2p} \to \{0,1\} \). Therefore, similar to Chan (1990) and Carrasco (2002), we can write \( E \left( s_{t-j} / F_{t-j}^{-1} \right) = \delta(c^*) \) and \( E \left( s_{t-j}^2 / F_{t-j}^{-1} \right) = \delta(c^*) \) for \( j \geq 0 \), note that \( c^* \) denotes the threshold parameter and \( \delta(c^*) \in (0,1) \). Using this results above,

\[
E \left( \Phi_{t-1}^{*} W_{k=2}^{m+1} (\Phi_{t-1}^{*})' / F_{t}^{-m-1} \right) = \Phi_{1} W_{k=2}^{m+1} \Phi_{1}' \delta(c^*) + \Phi_{2} W_{k=2}^{m+1} \Phi_{2}' (1 - \delta(c^*))
\]

Substituting, the result given above in (29),

\[
E \left( \Phi_{t-1}^{*} W_{k=2}^{m+1} (\Phi_{t-1}^{*})' / F_{t}^{-m-1} \right) = \Phi_{1} \left[ E \left( W_{k=2}^{m+1} / F_{t}^{-m-1} \right) \right] \Phi_{1}' \delta(c^*)
+ \Phi_{2} \left[ E \left( W_{k=2}^{m+1} / F_{t}^{-m-1} \right) \right] \Phi_{2}' (1 - \delta(c^*))
\]

Similar to Pham (1985), applying \text{vech} \ operator above, it is straightforward to show that equals to,

\[
\text{vech} \ E \left( \Phi_{t-1}^{*} W_{k=2}^{m+1} (\Phi_{t-1}^{*})' / F_{t}^{-m-1} \right) = \text{vech} \ E \left( W_{k=1}^{m+1} / F_{t}^{-m-1} \right)
= \left[ (\Phi_{1}' \otimes \Phi_{1}') \delta(c^*) + (\Phi_{2}' \otimes \Phi_{2}') (1 - \delta(c^*)) \right] \text{vech} \ E \left( W_{k=2}^{m+1} / F_{t}^{-m-1} \right)
\]

By repeated substitution of the former equation, we have,

\[
\text{vech} \ E \left( W_{k=1}^{m+1} / F_{t}^{-m-1} \right) = \left[ (\Phi_{1}' \otimes \Phi_{1}') \delta(c^*) + (\Phi_{2}' \otimes \Phi_{2}') (1 - \delta(c^*)) \right] \text{vech} \ E \left( W_{k=m+1}^{m+1} / F_{t}^{-m-1} \right)
\]
Next, as in Pham (1985), it can be showed that the right-hand side of the equation (28) is bounded by,

\[
E \left( \left\| A_{k=1}^{m+1} \right\|^2 / F_{t-m-1} \right) = Tr \left( E \left( W_{k=m+1}^{m+1} / F_{t-m-1} \right) \right) < \left| \text{vech} \ E \left( W_{k=m+1}^{m+1} / F_{t-m-1} \right) \right|
\]

Consequently, combining the last two result, we can deduce that,

\[
E \left( \left\| A_{k=1}^{m+1} \right\|^2 / F_{t-m-1} \right) < \left| \left( \Phi_1' \otimes \Phi_1' \right) \mathbf{\delta} \left( c^* \right) + \left( \Phi_2' \otimes \Phi_2' \right) \left( 1 - \mathbf{\delta} \left( c^* \right) \right) \right|^m \cdot \left| \text{vech} \ E \left( W_{k=m+1}^{m+1} / F_{t-m-1} \right) \right|
\]

Firstly, it is straightforward to show that \( \left| \text{vech} \ E \left( W_{k=m+1}^{m+1} / F_{t-m-1} \right) \right| \leq M_1 < \infty \), where \( M_1 \) is a nonnegative \( \left( \frac{1}{2} (p+1) \times 1 \right) \) constant vector depending of the coefficients of the matrices \( \Phi_1 \) and \( \Phi_2 \). From Lütkepohl (1996), we know that Assumption (2b) implies that \( \rho \left( \Phi_i' \otimes \Phi_i' \right) < 1 \) and \( \left( \Phi_i' \otimes \Phi_i' \right)^m \rightarrow 0 \) as \( m \rightarrow \infty \), for \( i = 1, 2 \), one obtains,

\[
E \left( \left\| A_{k=1}^{m+1} \right\|^2 / F_{t-m-1} \right) < \max_{i=1,2} \rho \left( \Phi_i' \otimes \Phi_i' \right)^m \cdot M_1
\]

Finally by Tower property, we have,

\[
\left( E \left( \left\| A_{k=1}^{m+1} \right\|^2 \right) \right)^\frac{1}{2} = \left( E \left( E \left( \left\| A_{k=1}^{m+1} \right\|^2 / F_{t-m-1} \right) \right) \right)^\frac{1}{2} < E \left( \max_{i=1,2} \rho \left( \Phi_i' \otimes \Phi_i' \right)^m \cdot M_1 \right)
\]

Therefore, we obtain the desired result,

\[
\left\| A_{k=1}^{m+1} \right\|_2 < \max_{i=1,2} \rho \left( \Phi_i' \otimes \Phi_i' \right)^m \cdot M_1 \rightarrow 0 \text{ as } m \rightarrow \infty
\]

Similar arguments lead us to prove that,

\[
\left\| \hat{A}_{j=1}^{j-1} \left( \Phi_1 - \Phi_2 \right) \Lambda_{k=j+1}^{k} \right\|_2 < \max_{i=1,2} \rho \left( \Phi_i' \otimes \Phi_i' \right)^\frac{1}{2} \left( m+j-2 \right) \cdot M_2 \rightarrow 0 \text{ as } m \rightarrow \infty
\]

The result is based on a double application of Cauchy-Schwartz’s inequality and discontinuity Assumption given in (2a). Note that, \( M_2 \) is a nonnegative \( \left( \frac{1}{2} (p+1) \times 1 \right) \) constant vector depending \( M_1 \) and \( \left\| \Phi_1 - \Phi_2 \right\|_2 \).

Applying the results of Lemma A2.2. in equation (25), we obtain,

\[
\left\| V_t - V_t^m \right\|_2 \leq C_1 \sum_{j=1}^{m} \sum_{k=1}^{m} \rho \left( m+j-2 \right)^{\frac{1}{2}} \cdot M_2 \left\| V_{t-j} - V_{t-j}^m \right\|^\frac{1}{2} + \rho^\frac{1}{2} \cdot M_1 \left\| V_{t-m-1} \right\|_2
\]

implying that since \( \rho < 1 \), by Lemma A2.2., \( \left\| V_t - V_t^m \right\|_2 \rightarrow 0 \) geometrically as \( m \rightarrow \infty \) or bounded of order \( \left\| V_t - V_t^m \right\|_2 = O \left( \max_{i=1,2} \rho \left( \Phi_i' \otimes \Phi_i' \right)^m \right) \).
The final result follows from the result above and it is sufficient condition for $L_2-$NED of $(V_t, t \geq 1)$ on the underlying $\alpha-$mixing sequence $\{\eta_t\}$ as in Definition 4. The results is based on the following bounded condition taken from Theorem 10.12 in Davidson (1994),

$$\|V_t - E (V_t | F_{t+m}^{t+m})\|_2 \leq \|V_t - \tilde{V}_t^m\|_2$$

7 REFERENCES


